Methods for Priority Setting Among Clinical Preventive Services

Michael V. Maciosek, PhD, Ashley B. Coffield, MPA, J. Michael McGinnis, MD, MPP,
Jeffrey R. Harris, MD, MPH, M. Blake Caldwell, MD, Steven M. Teutsch, MD, MPH, David Atkins, MD, MPH,
Jordan H. Richland, MPA, MPH, Anne Haddix, PhD

Overview: Methods used to compare the value of clinical preventive services based on two criteria—clinically preventable burden (CPB) and cost effectiveness (CE)—are described. A companion article provides rankings of clinical preventive services and discusses its uses for decision-makers; this article focuses on the methods, challenges faced, and solutions. The authors considered all types of data essential to measuring CPB and CE for services recommended by the U.S. Preventive Services Task Force and developed methods essential to ensuring valid comparisons of different services' relative value.

Medical Subject Headings (MeSH): cost-benefit analysis; data collection; delivery of health care; economics; healthcare quality, access and evaluation; health priorities; preventive health services; preventive medicine; quality-adjusted life years (Am J Prev Med 2001;21(1):10–19) © 2001 American Journal of Preventive Medicine

Introduction

The challenge of setting priorities among clinical preventive services lies in deriving consistent estimates of services' value using disparate data. A valid ranking of preventive services requires that each service's underlying data be evaluated on the same basis. This article addresses how data issues were handled, the potential impact on the results, and how various methods affected a priority ranking of clinical preventive services. Coffield et al.,[1] a companion article in this issue, provides this ranking and other results, as well as the rationale for evaluating services' relative value and a discussion of how decision-makers may interpret and use the findings.

Many of these methods have not been previously addressed in the literature. Although this discussion is detailed, we were unable to include particulars on the 30 different preventive services assessed in this analysis or the methods that are specific to evaluating each service. For service-level detail, readers are referred to a technical report on the evidence and calculations used

to derive estimates for ranking the services.[2] The technical report provides the methodologic details for each particular service that may be required to replicate or improve on this work.

General Approach

As summarized in the companion article,[1] the Committee on Clinical Preventive Service Priorities (the Committee) outlined the approach used to set priorities among clinical preventive services, identified the set of services to be included, and selected the variables on which the ranking should be based. The Committee chose to evaluate services recommended for the general population by the U.S. Preventive Services Task Force (USPSTF)[3] based on clinically preventable burden (CPB) and cost effectiveness (CE). The Committee also chose a 9-point scale (ranging from 2 to 10) for ranking these services.

To develop estimates of the magnitude of CPB and CE that are comparable across services, we needed a common time period and a standard rule for identifying the population to be included in these estimates. Both CPB and CE were based on delivery of the service to a 1-year, U.S. birth cohort over the age range the service is recommended by the USPSTF and at the recommended frequency of delivery. We did not include burden of disease prevented or costs of service delivery for populations continuing to receive the service after the recommended age unless it is part of disease management resulting from positive screening results.

Differences in the recommended frequency of delivery across services required that the conceptual approach have a time component. In fact, the level of effectiveness found in studies often reflects these recommended delivery intervals. When estimating CE, the frequency of delivery over the recommended age range defined the total costs of delivery necessary to achieve the level of effectiveness found in studies.

It was not possible to explicitly model the course of disease over time for each of the dozens of conditions covered by these 30 services. In most cases, we approximated the incidence and costs of disease that would be observed in a U.S. birth cohort over the age range the service is recommended by using the incidence and costs of disease observed in the age range recommended for the service in the most recent calendar year. This method yielded estimates that included approximately the same number of person-years (total years, and by age, gender, race, and ethnicity) as would be found when following an actual birth cohort over time. The accuracy of this approximation varied by health condition. Demographic change, changes in the risk factors for the condition, and changes in competing risks can cause the experience of the current cross-section to be different from what might be expected of a birth cohort that is entering the recommended age for the preventive service today. For HIV, the experience of a recent cross-section may be a particularly poor approximation due to the changing nature of the epidemic and recent developments in treatment. Thus, we used current incidence rates and an estimate of the projected burden of HIV given current treatment technologies.[4]

See related Commentary on page 66.

Data Sources

Studies were obtained by searching MEDLINE and HealthSTAR. These searches were supplemented by citations obtained from experts and reviews of other sources, particularly the USPSTF's *Guide to Clinical Preventive Services*, second edition.[3] Only tabulated data from national databases and published estimates were used. Many types of data were needed (including, for example, incidence rates, effectiveness, and portion of the population at risk), and for each of these and other basic types of data, variations were necessary for each service. Some services required well over 100 separate data elements. Given the large number of data elements required, many data elements were available from only a few sources, and data quality varied significantly. As a result, all types of evidence were considered, and systematic exclusion criteria were not employed. We chose to seek the best estimates given the available data rather than set data quality standards that precluded the use of lower quality information when no other information was available.

Study quality and the applicability of available data to our analyses were considered in choosing estimates. If given a choice among estimates, the better candidate was chosen based on such factors as well-defined control groups, randomized trial, larger study population, and lack of selection bias. More recent estimates were preferred in cases where the estimate could be affected by recent trends in risk factors or technology. Although estimates from meta-analyses were preferable, these often lacked the reporting detail needed to assess the appropriateness of the underlying studies for our purposes. Thus, the individual studies were frequently used. If we had a choice among estimates and the best one was not obvious, we averaged all available estimates. If some were obviously poorer in quality, these estimates were excluded from the average. More complete searches were conducted for variables that appeared most important to our estimates of CPB and CE once initial estimates were generated. For some variables, experts were consulted to identify best-quality evidence and to provide ranges of estimates when data were conflicting or not found. On two occasions during the course of the study, subject-area experts from the Centers for Disease Control and Prevention reviewed the underlying data and the CPB and CE calculations.

Clinically Preventable Burden

Estimates of CPB based on observations of individuals receiving services are not available. Instead, CPB must be calculated from disparate sources. Conceptually, CPB is the burden addressed by the service multiplied by the effectiveness of the service. When combined in this way, effectiveness has a very specific definition: the percent reduction in burden addressed by the service. Burden addressed by the service ("addressable burden"), on the other hand, can vary in definition as long as it is paired with an appropriate estimate of effectiveness. For example, the burden addressed by cholesterol screening could be defined as: (1) all cardiovascular disease (CVD), (2) the portion of CVD caused by high cholesterol, or (3) all CVD (regardless of cause) that occurs in people with high cholesterol. The only requirement for a definition of addressable burden is that it includes the entire burden that can be prevented. Addressable burden could, in the extreme, include all-cause mortality if a suitable estimate of cholesterol screening's ability to reduce all-cause mortality were available. The broader the definition of addressable burden, the smaller the corresponding estimate of effectiveness (measured as percent of addressable burden prevented). No differences should be found in the absolute values of CPB for

cholesterol screening as long as effectiveness and burden correspond to one another.

In practice, there is usually very little choice in estimates of effectiveness. As a result, the available effectiveness data typically dictate the definition of addressable burden for each service. For cholesterol screening, available measures of effectiveness are from clinical trials that enrolled individuals with established high cholesterol; these trials do not distinguish CVD outcomes by cause (e.g., CVD events attributable to high cholesterol vs events attributable to high blood pressure). Therefore, the definition of addressable burden necessary to arrive at a valid measure of CPB is all CVD (regardless of cause) in people with high cholesterol (definition 3 above).

Measuring Burden

For services to be compared in a meaningful way, CPB must be measured in the same units for all services. Quality-adjusted life years (QALYs) are a recommended and increasingly common measure of health outcomes in cost-effectiveness analyses.[5,6] QALYs measure health status over time by adding the number of years of life at full health to years of life lived with illness or disability, where a year spent with illness or disability receives a lower weight than a year in full health.

Similarly, QALYs lost is a measure of the consequences of disease or injury in which the years of premature death (life years lost) are added to years spent with illness or disability. If a year spent in full health were valued at 1.0 and a year spent with a particular disease were valued at 0.60, for example, then a year of illness caused by the disease would be calculated at 0.40 QALYs lost (or 40% of a life year lost).

A small number of tables of disease- and condition-specific QALY weights[7–9] and disability weights[10,11] have been published, and each uses different methods for determining the weights. In these tables, regardless of method, all but the most severe and least severe conditions fall in the range of approximately 0.5 to 0.9 for short-duration illness or injury and 0.7 to 0.9 for chronic conditions. No published table of condition-specific weights includes all of the health conditions addressed by the set of clinical preventive services included in this analysis. For our base-case estimates, we adopted the midpoints of these ranges: 0.7 for short-term conditions and 0.8 for chronic conditions.

Compared to perfect health valued at 1.0, an illness or injury resulting in a QALY weight of 0.7 would indicate a quality of life lost of 0.3 QALYs per year $(1.0 - .7 = 0.3)$. Most weighting systems find that people without existing medical conditions have a quality of life of less than 1.0, usually about 0.9. A more precise measure of the quality of life lost to a medical condition with a weight of 0.7 would therefore be 0.2. However,

rather than re-scaling all years of life lost and all quality of life lost to a 0.9 scale, we evaluated years of life lost at 1.0 and estimated quality of life lost to morbidity by subtracting the QALY weight from 1.0. This approximation slightly overstates burden of disease across most services.

Due to higher rates of comorbid illness, older populations without the health condition of interest are likely to have a true quality of life weight further from 1.0. The use of 1.0 to approximate the true quality of life will cause the burden of disease estimates to be overstated for all fatal conditions regardless of the age of the target population. However, the approximation may cause artificial differences across age groups for morbidity, particularly for a few services where morbidity is substantial. Three services with a high portion of burden attributable morbidity in younger populations—screening services for newborns, chlamydia screening, and counseling on infant nutrition—may have low CPB estimates compared to all other services. For these services, the 1.0 is a better estimate of the true quality of life among persons without the condition of interest. The CPB for these services may be low on a relative basis because the CPB of other services is slightly overstated.

Exceptions to the QALY weights were made for those conditions clearly shown by the available information to be either more or less severe than for other conditions that fall within the ranges of 0.5 to 0.9 for short-duration illness and 0.7 to 0.9 for chronic conditions. Base-case estimates were assigned based on review of estimates from the published tables.[7–9] Other exceptions were made when a service's estimates of CPB and/or CE were particularly sensitive to the estimate of QALYs lost. These services primarily or exclusively addressed morbidity instead of mortality. For conditions addressed by these services, published scales were reviewed to identify more precise base-case estimates. Table 1 lists all the conditions for which alternate weights were used.

Duration of illness or injury was also an important consideration and logically connected to the QALY weight. In general, the severity of a condition will change over time. Ideally, the QALYs lost weight would be measured at frequent intervals over the duration of illness, and we would multiply the true incidence by duration of illness and the average QALYs lost weight. However, for many acute conditions, we relied on medical encounters (e.g., hospital inpatient stays, emergency department visits, and other outpatient visits) as approximations of incidence. The average severity of disease is likely to vary by type of medical encounter. When only estimates of emergency department visits or hospital inpatient stays were available, we presumed that only the more severe cases were represented and assigned a longer duration of illness than if outpatient estimates had been available. Thus, we varied the

Table 1. Exceptions for quality of life lost (base-case estimates only)

Condition	QALY weight
Drowning vegetative states	.80
Head injuries	.35
Hip fractures	
Short-term recovery	.35
Long-term recovery	.25
Mental retardation	
Mild	.30
Moderate	.50
Severe	.80
Infertility	.12
Stroke	.40
Long-term effects of severe iron-deficiency anemia	.10
Tooth loss (enough loss to require dentures)	.10
Long-term effects of untreated amblyopia in childhood	.02
Vision impairment (older adults)	.10
Hearing impairment (older adults)	.10

QALY, quality-adjusted life year.

average duration of illness according to the type of incidence data available, rather than varying our standard QALY weight.

Actual data on the duration of morbidity for acute conditions are rare. Estimates of restricted-activity days (including, but not limited to days spent in bed, days lost from work, and days lost from school) have been tabulated from the National Health Interview Survey.[12] For other acute conditions, we assigned duration based on the perceived severity and likelihood of extended disability relative to the conditions for which restricted-activity days have been tabulated.

For chronic conditions, it is generally possible to identify more-accurate estimates of incidence and duration. Individual studies frequently provide estimates of annual incidence, though occasionally only point prevalence estimates could be identified in individual studies. For many chronic conditions, we used incidence rates and average duration estimates for established market economies from the Global Burden of Disease Study.[13] For a few conditions, duration was assigned based on life expectancy at the estimated average age of onset.

Most often we obtained mortality data from U.S. death tables,[14] but in many cases estimates from other sources, such as studies describing risk-factor attributable deaths, provided more appropriate estimates. For mortality, the cross-sectional data provided estimates of the average age at death. We estimated the average number of years lost at death by subtracting the average age at death from life expectancy at that age. We used life expectancy for the general population for all conditions.

We did not value QALYs realized in the long term

differently than QALYs realized in the short term, but QALYs are discounted for CE. Decision-makers will need to decide if they want to place greater weight on services with immediate or longer-term benefits.

The Impact of Morbidity Approximations on CPB

The number of approximations necessary to estimate morbidity prevented across a broad range of clinical preventive services may raise concerns about the accuracy of estimates of CPB. However, mortality incidence is the most important variable for most services. In 18 of 30 services, morbidity is <25% of CPB. In two services, morbidity is 50% and 75%, respectively, of the CPB estimate. For the remaining 10 services, the morbidity estimates are 100%. Only for these services does the choice of QALY weight and duration have a potential impact on the CPB score, which is extremely unlikely to change by more than 1 as a result of changes to QALY weights or duration of morbidity. CE estimates are even less sensitive to changes in the morbidity components.

With the exception of counseling about the risks and benefits of hormone replacement therapy, the services recommended by the USPSTF have minimal potential harms relative to the benefits. For this reason, harms are not included in estimates of CPB.

Measuring Effectiveness

As noted above, CPB estimates are based on specific definitions of effectiveness. Like addressable burden, effectiveness is not readily available and, thus, must be calculated in most cases. The actual components involved in measuring effectiveness vary from service to service.

A general conceptual equation is:

Effectiveness = (percent who would accept the preventive service once offered) × (sensitivity of screening or assessment) × (adherence with follow-up treatment or advice to change behavior) × (effectiveness of prevention, treatment, or behavior change)

In all cases, careful consideration of these components and the extent to which adherence is accounted for is necessary. For example, estimates of effectiveness in clinical trials with disease or mortality endpoints typically involve incomplete adherence with treatment. However, because the population includes study volunteers who are likely to receive closer follow-up than would be found in usual care, adherence with treatment may be overstated. In addition, the magnitude of adherence reported depends on how the data were analyzed. Estimates based on intention-to-treat (ITT) analysis[15] generally reflect more complete estimates of

adherence than estimates that include only those for whom follow-up measures were available.

In addition to adherence, the most common type of missing effectiveness information is effectiveness of services in preventing nonfatal cases. In most cases, available estimates of effectiveness in preventing mortality were applied to estimates of morbidity. The adequacy of this approximation depends, in part, on the mechanism of prevention. If deaths were averted as a result of preventing the incidence of disease or injury, the approximation should be adequate. If deaths were averted as a result of reducing the case-fatality rate, the adequacy of the approximation will vary. Among services that prevent both mortality and morbidity, morbidity is a large part of addressable burden in only two services—newborn screening and counseling on sexually transmitted disease (STD) risks. Therefore, the CPB estimate is generally not affected by using what is known about effectiveness in reducing mortality for both mortality and morbidity.

Effectiveness of Counseling Services

One of the most challenging tasks was estimating the level of effectiveness of counseling services, where effectiveness essentially reflects two elements: efficacy of risk reduction (or the efficacy of behavior change) and adherence with clinician advice to change behavior. The USPSTF assigned two separate strength-of-recommendation grades to the counseling services: one for the efficacy of risk reduction accomplished through behavior change and another for adherence with clinician advice to modify behavior. A counseling service, therefore, might receive grades such as A/C, where "A" corresponds to strong evidence of the efficacy of behavior change and "C" to insufficient evidence of adherence with clinician advice. The majority of counseling services received strength-of-recommendation grades of "C" for adherence with clinician advice. Despite insufficient or weak evidence on adherence with clinician advice, most such counseling services are recommended by the USPSTF if the grade for efficacy of risk reduction were an "A" or "B" by USPSTF criteria.

For counseling services, we interpreted the USPSTF recommendations for brief clinician advice to imply counseling lasting ≤5 minutes for the average patient. No recommendations on the exact frequency of counseling are provided. When assigning adherence estimates, we considered every-other-year delivery to be feasible. However, the available data are not sufficient to determine how the frequency of delivery affects adherence.

A critical consideration was the impact of repeated advice over the age range recommended by the USPSTF. The vast majority of literature measures effectiveness of one-time advice (defined as a single counseling session or several sessions in a short period of time). Without periodic counseling, adherence may naturally decline over time. With periodic counseling, adherence may still decline over time or be maintained at short-term levels. It may increase if behavior change also occurs among individuals who were initially unreceptive to advice, but are at a more receptive stage of change at the time of subsequent counseling.

We did not apply formal quality-of-study assessment tools for studies of adherence with brief clinician advice. However, by any reasonable standard, virtually all of these studies would be considered limited or extremely limited in terms of study quality or applicability of study results to calculations of CPB. Limitations include but are not limited to: lack of a control group (in some studies, people receiving brief counseling constituted the control group); large potential for self-selection of study volunteers; small sample sizes; large losses to follow-up; very short follow-up; study populations at greater risk than the general population; measures of adherence not reported (e.g., results may be reported as changes in fiber content of diet); intensive counseling rather than brief counseling; and self-reported adherence. We divided studies into "less-limited" and "more-limited" categories on a qualitative basis. The large differences among studies, the limitations, and the small number of studies made systematic review impossible. A qualitative review of available studies is summarized here. Further detail and a complete list of references are available in the detailed report on the evidence and calculations.[2]

In general, the estimates we report have been adapted from those reported by the authors. We used the studies' control groups to calculate adherence as the percent of individuals who would not have otherwise changed their behavior. When possible, we report the authors' estimates based on ITT analysis or our own approximations of ITT analysis.

Counseling adult smokers to quit has the strongest body of evidence on adherence for any single counseling service. The USPSF gave this service A/B grades. Most counseling services received A/C grades. A meta-analysis indicated that 2.5% of smokers who would not have otherwise quit did so following ≤3 minutes of clinician advice (based on short-term follow-up).[16] Another review calculated the expected 1-year quit rate considering the possibility that some smokers would attend more extensive counseling and/or use nicotine replacement therapy after brief counseling.[17] The results indicate a 3.4% success rate among those who would have otherwise remained smokers at 1 year. Two studies of adult smoking-cessation counseling provide the only direct evidence of the effectiveness of repeated counseling delivered over several years.[18,19] These studies show approximately 30% reductions relative to controls. However, both studies involved individuals at high risk for heart disease and date to the early 1970s. We calculated a rough estimate of adherence with

advice delivered repeatedly over time (27%) based on the 1-year adherence estimates cited above, reported 1-year delivery rates for brief counseling, and the reported changes in adherence over time from the counseled and control groups in the two long-term studies. Using the results from sensitivity analysis, we derived a range for adherence of 17% to 39%.

Five brief counseling services for adults lacked adequate information for estimating adherence with clinician advice. No adherence estimates were found for counseling on oral health practices, counseling dependent problem drinkers to enroll in a clinical or community treatment program, and counseling individuals (excluding young children) on injury risks. Only a small number of "more-limited" studies were found for counseling on dietary patterns and counseling on the risks of STDs. (Studies that included STD counseling following HIV testing were excluded.) Studies of counseling on the risk of STDs seem to indicate that some positive level of adherence is plausible among high-risk groups, but contradictory results and lack of control groups for brief counseling make stronger conclusions impossible. For a sixth service, physical activity counseling, positive adherence of <10% is indicated by a small set of "less-limited" studies.

We assigned a range of 0% to 5% adherence for these six counseling services and used a midpoint estimate of 2.5% adherence. Figure 2 and Figure 3 in Coffield et al.,[1] in this issue, show these services with a minimum of 1% adherence in order to depict differences between the services at low levels of adherence. At 0% adherence, CPB is zero and the CE ratio is infinite.

We assigned counseling parents on childhood injury prevention, counseling nondependent problem drinkers to modify their alcohol use, and counseling to promote daily intake of calcium/vitamin D and folic acid among women a midpoint adherence estimate of 12.5%. For each of these services, a small cluster of estimates from studies of varying quality and applicability range from 5% to 20% within a broader range of 0% to 60%.

The two counseling services for adolescents (counseling to avoid tobacco use and counseling to avoid alcohol and drugs) both have relatively extensive literatures on one-time classroom interventions, but no evidence on the effectiveness of clinician counseling. The evidence on one-time classroom programs, such as Drug Abuse Resistance Education (DARE), indicates that the behavior of small proportions of students may be affected in the short term but not in the long term. Other studies indicate that programs with multiple sessions over several grades may have short-term success rates and possible lower levels of long-term success rates. In all cases, we cannot rule out the possibility of zero adherence (short term and long term), given that these studies do not provide direct evidence on adher-

ence with clinician advice. We assigned ranges of 0% to 2% and used 1% as our midpoint estimate for long-term adherence with brief counseling for these services. To show the difference between the services at lower levels of adherence, we used 0.5% as our minimum level of adherence (Figure 2 and Figure 3 in Coffield et al.[1] in this issue). As used in CPB calculations, 1% adherence refers to 1% of the 35% (roughly) of adolescents who would otherwise become adult smokers for some period of time (or 0.35% of all adolescents).

It is possible that those who do adhere with clinician advice have a higher or lower risk of illness or injury resulting from the targeted behaviors than the general population. For example, individuals with a family history of heart disease may be more likely to follow clinician advice to increase physical activity levels. When calculating CPB, it is implicitly assumed that the proportion of people who adhere are at average risk. No data are available to indicate whether or not this assumption is valid.

Adjusting Burden for Delivery Rate

For each service, CPB was estimated independently of current delivery of the service to indicate the service's total value, rather than the value of improving delivery over its current level for the U.S. population. For some services, the addressable burden is estimated by multiplying incidence rates in populations not receiving the preventive service by the size of the target population; therefore, no adjustment for the current delivery rate was necessary. Historical incidence rates were used for childhood immunizations and the Td booster. These estimates may overstate burden of disease in the absence of vaccines because better disease control measures and improved treatment might lead to lower incidence and lower case-fatality rates than were observed prior to vaccine development. Some services were not adjusted because the current burden of disease differs from burden of disease at zero delivery of the preventive service by <1%; this is due to both low current delivery rates and low levels of service effectiveness.

For other services, burden at a zero delivery rate was estimated as $BD_{ZDR} = BD_{CDR} + [1 - (CDR \times Eff)]$, which is derived from substituting equation (2) below into equation (1) and algebraic manipulation:

$$BD_{ZDR} = BD_{CDR} + BDP_{CDR} \qquad (1)$$

where BD_{ZDR} = projected Burden of Disease with zero delivery rate of the clinical preventive service; BD_{CDR} = currently observed Burden of Disease at the current delivery rate of the preventive service; and BDP_{CDR} = Burden of Disease prevented at the current delivery rate.

$$BDP_{CDR} = BD_{ZDR} \times CDR \times Eff \qquad (2)$$

where CDR and Eff = Current Delivery Rate and Effectiveness, respectively, of the preventive service.

It is important to note that all of these adjustments for current delivery rates are limited exclusively to the services being analyzed. For tobacco cessation counseling, for example, we only adjusted the CPB estimate for the absence of counseling, not for the absence of all other interventions and information that help smokers quit.

Cost Effectiveness

Scores for CE were based on existing CE studies when available. For purposes of comparability, we searched for studies that were conducted following the "reference-case" methods recommended by the Panel on Cost Effectiveness in Health and Medicine (PCEHM)[5] and the methods described above for estimating CPB. We sought studies that had taken the societal perspective, used a 3% discount rate, measured health outcomes as QALYs, and only included time costs associated with receiving the service.

Because the limited number of CE studies that exist for clinical preventive services did not meet all the criteria necessary for comparability and were not consistent with the CPB estimates, we developed a systematic method for adjusting CE ratios.[20] For each service, we selected the study that most closely matched the service description and target population recommended by the USPSTF and followed the PCEHM recommendations. The study must also have reported enough detail on costs and outcomes to permit adjustments of results to improve comparability.

We converted costs to 1995 U.S. dollars using the appropriate consumer price index.[21] Costs and outcomes were recalculated using a 3% discount rate. Where appropriate, time costs associated with receiving the clinical preventive service were added. We assumed an average cost of $25 for patient time and travel expenses for services that required the patient to schedule a separate visit. We added 25% of that cost for services that could be added on to an already scheduled visit. No patient time and travel costs were added for brief counseling or simple screening services. Productivity losses associated with morbidity and premature mortality were excluded. Medical costs that are not related to the disease or injury addressed by the service were also excluded. CE ratios were calculated using a "no-service" comparator. A more complete explanation of the adjustment process is made available elsewhere.[21]

When published estimates were not available, we used a simplified framework to estimate CE ratios. We followed the PCEHM guidelines and used data collected from CPB and other sources. We used service costs and the cost of health conditions from the existing literature. Costs averted were obtained by applying the effectiveness of the service (from the CPB evidence review) to estimates of the costs of the health condition. Costs were converted to 1995 U.S. dollars and discounted at 3%. Because these estimates lack the precision of a formal analysis, extensive sensitivity analyses were conducted.[5,6]

Most counseling services lacked published CE studies. Thus, we used an average of $5 of clinician time for patients receiving a brief counseling session. Some patients will require (or accept) only a few seconds of counseling, and other patients will seek more extensive information and advice. Five dollars would cover an average of 2 minutes per patient, with an hourly rate of $150 to cover salary and overhead. We assigned from $5 to $25 for the implementation of formal screening tools, such as a written risk assessment questionnaire or extensive history taking. In sensitivity analysis, CE ratios for counseling services were estimated across a range of delivery frequencies, generally from annual delivery to delivery every 3 years and across the range of adherence rates defined in the CPB estimates. CE scores for counseling services are based on the costs of delivery once every 2 years and the midpoints of adherence estimates.

CE ratios are reported as the cost per QALY saved. The range used for each scoring quintile is reported in Table 1 of the accompanying article.[1]

Calculating the Ranking

The CPB and CE "base-case" estimates were sorted and separated into quintiles (Table 2 of accompanying article[1]). Each service was assigned a CPB score and a CE score, each ranging from 1 to 5 (according to quintile), with 5 being the best possible score. Services with the highest CPB were assigned a CPB score of 5, and services with the lowest CE ratios were assigned a CE score of 5. These scores were added to give services a total score on a possible scale of 2 to 10 (integer values only). In addition to total scores, individual CPB and CE scores are presented for each service to allow decision-makers to consider each criterion separately.

Because of the uncertainty in CPB and CE estimates, this ordinal scoring method provides an indication of the relative value of services within broad categories rather than ranking all services based on CPB and CE point estimates. However, addition of ordinal scales may result in spurious total scores.[22] Thus, it is possible that investing in improving delivery of a service with a CPB-CE score combination of 3-4 is a worse decision than investing in service with a score combination of 4-2. This is because the loss in the CE score of 2 (moving from the fourth to the second quintile in the distribution of CE ratios) may be less important than increasing the score by 1 in the CPB scale (from the third to the fourth quintile of the CPB scale).

Sensitivity Analysis

Sensitivity of scores to changes in the values of the underlying CPB and CE variables was assessed for all services. We reported ranges of CPB and CE in Figure 2 and Figure 3 of the companion article in this issue, Coffield et al.,[1] using the methods described below. We also identified the services with scores that are particularly sensitive to underlying data estimates.

CPB and CE Estimates

For all but the most certain variables, we defined plausible ranges for each data point using the base case as the midpoint. We then conducted two series of single-variable and multiple-variable sensitivity analyses to identify the three variables that collectively cause the largest change in CPB and CE estimates. We call these the "most influential" variables. We calculated a range of CPB and CE estimates based on the highest and lowest estimates for the set of the "most influential" variables.

Because the CPB and CE estimates are derived from variables that may be related to one another, two factors must be considered. First, although changes in the value of an individual variable may not substantially influence the CPB or CE, changes in all related variables may substantially influence the estimates. Second, because the variable is related to others, it is likely that an under- or over-estimate of the value of one variable may result from systematic bias in the measurement of all related variables. For example, if there are ten or more causes of mortality in an estimate of QALYs lost for a particular service (e.g., tobacco cessation counseling), no single component of mortality is likely to have a substantial impact on either the CPB or the CE estimates. However, the sources of mortality data or the population-attributable fractions may be systematically biased, causing a substantial adverse impact on the CPB or CE estimate.

To guard against systematic error in CE estimates, we ran a second series of sensitivity analyses in which we grouped related variables (e.g., total discounted QALYs lost, total discounted intervention costs, total discounted costs of illness, and total adherence) and treated the group as an individual variable. If a group variable was more influential than any one of the three "most influential" variables from the first series of sensitivity analyses, it was substituted for the least influential of the most influential variables. For CPB, we analyzed the potential for systematic error in the components of QALYs lost (mortality incidence, average life expectancy, morbidity incidence, morbidity duration, and morbidity QALY weight), rather than total QALYs lost.

CE estimates based on adjustments to published estimates required a different strategy because we did not reconstruct the model from each data point; studies virtually never reported all data points. For adjusted CE ratios from published studies, we report ranges based on the two most influential variables among total discounted costs, total discounted savings, and total discounted QALYs saved. We generally used a 25% variation in these aggregate variables, but used 10% to 50% variation to reflect the reliability of the main data source(s).

Sensitivity of Scores

A change in the score of one service will always affect the score of another service, as one service "bumps" another from its quintile. However, a change in the CPB or CE estimate of one service does not always change its score. In this way, the scoring system protects a service's ranking and the rankings of other services from small to moderate changes in the estimates. Large changes in estimates—either from changes to a single very influential variable or changes to several less influential variables—may be possible, and numerous scenarios exist in which the score of one service (and at the same time another service) will change. Most services could experience a small change in score; thus, differences in total scores of only 1 must be treated with caution.

One or two influential variables that are particularly uncertain may cause CPB, CE, or both, to be particularly sensitive for some services. In a subset of these cases, the plausible changes in the variables may cause changes in total scores of ≥2, thereby affecting the scores of more than one other service. We have indicated these services in Table 2 of the accompanying article.[1] In Figures 2 and 3 of the same article, these services generally have wide ranges of uncertainty in the underlying CPB and CE estimates. In most of these cases, the uncertainty is due either to uncertain estimates of adherence in counseling services or to uncertain estimates of QALY weights or duration of chronic conditions in services that do not prevent deaths.

For two of these services—screening for problem drinking and counseling and ocular prophylaxis for newborns—the large degree of uncertainty comes from very unstable CE ratios. In both, the costs of prevention and potential cost savings are large, but the discounted net costs are small. Small changes in either the costs of prevention or costs saved produce large percentage changes in net costs. Uncertain estimates of adherence contribute to the instability of the CE ratio of screening for problem drinking, and a very small denominator (discounted QALYs saved) contributes to the instability of the CE ratio for ocular prophylaxis. Due to the instability, these CE ratios could plausibly take on values consistent with any of the CE quintiles, and any point estimate of the CE ratio is nearly arbitrary.

Therefore, these services have been assigned the middle CE score of 3.

Discussion

A number of key findings emerged from this study and have been discussed; some of the more important ones are summarized here. First, to estimate CPB for a birth cohort, the definition of addressable burden is determined by available effectiveness data. Second, data needs were larger than expected, primarily due to the number of health conditions addressed by each service, but also due to multiple components of effectiveness, including adherence at more than one level. We found that establishing the magnitude of effectiveness is more difficult than establishing effectiveness. Third, all variables necessary to calculate CPB are not equally important; QALY weights and the duration of nonfatal disease and injury are relatively insignificant to the results when making comparisons across a wide range of preventive services, most of which prevent deaths.

In addition, adherence is the most important variable determining the ranking of counseling services, which make up almost half of the services in the analysis. (Services commonly referred to as "chemoprophylaxis," such as folic acid and calcium/vitamin D intake, fall neatly into the category of counseling when the components necessary to measure the magnitude of effectiveness—efficacy of the "behavior" and adherence with clinician advice—are considered.) Yet the literature available to estimate adherence for counseling services is small and limited in quality, particularly for brief counseling delivered repeatedly over time. Very low levels of adherence (e.g., <1%) do not indicate that a service is ineffective. In fact, very small levels of adherence can have a substantial effect on population health, and small absolute changes in adherence (from 1% to 2%) can have a large impact on estimates of CPB and CE. Counseling adolescents about tobacco use was assigned the lowest adherence, which some might equate to zero, yet ranked at the top of the list of services. We strongly encourage more study of patient adherence with repeated, brief clinician advice in order to reduce uncertainty in decision-making. However, as previously reported, extremely large trials may be required for some counseling services to demonstrate statistically significant improvements in health outcomes at the low levels of adherence that produce cost-effective counseling.[23] In some cases, evidence of the effectiveness of clinician counseling may only be possible from modeling; waiting for definitive evidence is tantamount to never improving delivery of these services.

Some of the limitations to the methods used to prioritize services are addressed in the accompanying article[1] but are briefly reviewed here. For example, not all considerations important for decision-making, such as ease of increasing delivery rates and the distribution of health benefits to high-risk groups, are included in the ranking. Comparing the rankings to national delivery rates also provides an adequate indication of what national priorities should be, but local decision-makers will need to consider their own current delivery rates. In addition, the methods assume that all people are seen by clinicians on a regular basis, which inflates the estimates. This may be insignificant in a comparison of services' value, but important when considering what delivery rates to improve first. As mentioned in this article, adding scores for CPB and CE may result in spurious scores; thus, decision-makers may want to consider each criterion separately. Finally, these methods do not eliminate uncertainty about services' relative value, but uncertainty is definitely minimized by the scoring system.

Feasible improvements to some of the methods include using our own cross-tabulations from national data sets rather than relying on published estimates. At the outset of this analysis, the vast number and types of variables needed were not predicted, but researchers will be better prepared for future analyses. Expert panels, rather than individual subject experts, could also be used to estimate missing data. Finally, we could conduct more precise modeling techniques and Monte Carlo simulations in sensitivity analysis. These improvements would add marginal precision. The most uncertain estimates are due to uncertain data, and expert panels and other improvements may not decrease the level of uncertainty.

Researchers are urged to pay careful attention to the components necessary to measure CPB. Careful collection and reporting of patient adherence at all levels, the training of clinicians involved in trials, and frequency and duration of morbidity observed in addition to mortality would result in more precise syntheses of CPB and aid CE analyses. Specifically for estimates of adherence, more attention to selectivity bias of volunteers, reporting loss to follow-up and ITT analysis, and contamination of control groups would be helpful. The time path of quality-of-life changes for health conditions (such as hearing and vision loss, infertility, and hip fractures) is also needed, as is better incidence data for injuries, chlamydia infection, and other STDs. Needed cost estimates include the costs of brief counseling and the costs of brief screening instruments (such as the CAGE screen[24] recommended by the USPSTF). Data on the efficacy of behavior changes (e.g., becoming physically active or, in particular, changing dietary practices) could also be strengthened.

To determine where to focus efforts to improve the delivery of preventive services, decision-makers need data syntheses that move beyond whether or not services are effective. These data syntheses must include as much of the available data as reasonably possible, summarize these numerous data points as meaningful

measures of each service's value, evaluate each service in a consistent manner to improve the validity of comparisons across services, and present the results in a useful format that does not grossly overstate estimates' precision. To our knowledge, this is the first work to address these needs across a broad range of recommended preventive services. Alternative methods could have been chosen; those used here should serve as a strong foundation on which to build future efforts. Future analyses might narrow their focus to specific populations or widen their scope to include community preventive services.

References

1. Coffield AB, Maciosek MV, McGinnis MJ, et al. Priorities among recommended clinical preventive services. Am J Prev Med 2001;21(1):1–9.
2. Partnership for Prevention. Evidence and calculations: clinically preventable burden and cost effectiveness for preventive services recommended by the U.S. Preventive Services Task Force. Washington, DC: Partnership for Prevention, 2001. Available at: www.prevent.org.
3. U.S. Preventive Services Task Force. Guide to clinical preventive services, 2nd ed. Baltimore, MD: Williams & Wilkins, 1996.
4. Holtgrave DR, Pinkerton SD. Updates of cost of illness and quality of life estimates for use in economic evaluations of HIV prevention programs. J Acquir Immune Defic Syndr Hum Retrovirol 1997;16(1):54–62.
5. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.
6. Drummond MF, O'Brian BO, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes, 2nd ed. New York: Oxford University Press, 1997.
7. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making 1993;13(2):89–102.
8. Gold MR, Franks P, McCoy KI, Fryback DG. Toward consistency in cost-utility analyses: using national measures to create condition-specific values. Med Care 1998;36(6):778–92.
9. Mittmann N, Trakas K, Risebrough N, Liu BA. Utility scores for chronic conditions in a community-dwelling population. Pharmacoeconomics 1999;15(4):369–76.
10. Murray C, Lopez A. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Global burden of disease and injury series. Vol. I. Geneva, Switzerland: World Health Organization, 1996.
11. Mathers C, Vos T, Stevenson C. The burden of disease and injury in Australia (AIHW Cat. No. EHE 17). Canberra, Australia: Australian Institute of Health and Welfare (AIHW), 1999.
12. Adams PF, Hendershot GE, Marano MA. Current estimates from the National Health Interview Survey, 1996, National Center for Health Statistics. Vital Health Stat 1999;10(200).
13. Murray C, Lopez A. Global health statistics: a compendium of incidence, prevalence, and mortality estimates for over 200 conditions. Global burden of disease and injury series. Vol. II. Geneva, Switzerland: World Health Organization, 1996.
14. Centers for Disease Control and Prevention. CDC WONDER: compressed mortality/population data request screen, 1968–1997. Available at: http://wonder.cdc.gov/mortJ.shtml. Accessed November 15, 2000.
15. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999;319(7211):670–4.
16. U.S. Department of Health and Human Services. AHCPR clinical practice guideline No. 18: smoking cessation. Washington, DC: Agency for Health Care Policy and Research, 1996 (Pub. No. 96-0692).
17. Cromwell J, Bartosch WJ, Fiore MC, Hasselblad V, Baker T. Cost-effectiveness of the clinical practice recommendations in the AHCPR guideline for smoking cessation. Agency for Health Care Policy and Research (AHCPR). JAMA 1997;278(21):1759–66.
18. Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial: risk factor changes and mortality results. JAMA 1982;248(12):1465–77.
19. Hjermann I, Velve Byre K, Holme I, Leren P. Effect of diet and smoking intervention on the incidence of coronary heart disease: report from the Oslo Study Group of a randomised trial in healthy men. Lancet 1981;2(8259):1303–10.
20. Carande-Kulis VG, Maciosek MV, Briss PA, et al. Methods for systematic reviews of economic evaluations for the guide to community preventive services. Am J Prev Med 2000;18(suppl 1):75–91.
21. U.S. Bureau of the Census. Statistical abstract of the United States, 1999. 119th ed. Washington, DC: U.S. Bureau of the Census, 1999.
22. McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. New York: Oxford University Press, 1987.
23. Downs SM, Klein JD. Clinical preventive services efficacy and adolescents' risky behaviors. Arch Pediatr Adolesc Med 1995;149(4):374–9.
24. Mayfield D, McLeod G. The CAGE questionnaire: validation of a new alcohol screening instrument. Am J Psychiatry 1974;131(10):1121–3.