Cancer Outcomes Research


If the central business of cancer outcomes research is the pursuit of information relevant to a range of decision makers (1), i.e., patients, families, providers, payers, regulators, standards setters, and researchers, several questions quickly arise. What is the nature and scientific quality of the information currently being produced? How do we enhance the rigor and relevance of cancer outcomes research, with a concern not only for the methodological and empirical foundations but also its potential and actual contributions to decision making? What is the research agenda to carry us forward?

The preceding 10 papers of this Monograph sought to address the first question above—the quality of the information—through empirically grounded reviews and evaluations of the published literature. This final paper, authored by staff at the National Cancer Institute (NCI), examines aspects of the remaining questions. It identifies the elements of a research agenda intended to generate better scientific products and information to enhance the quality of cancer care decision making and ultimately the quality of cancer care.

We have said that the purpose of cancer outcomes research is to describe, interpret, and predict the impact of interventions and also other influences on “final” outcomes that matter to decision makers (1). Such outcomes include not only survival and disease-free survival but also important nonbiomedical, patient-reported outcomes such as health-related quality of life (HRQOL), patient perceptions of and satisfaction with health care, and the economic burden attributable to cancer and its interventions. Final outcomes are distinguished from both intermediate outcomes (e.g., appropriate cancer screening) and clinical outcomes (e.g., delay in tumor progression), which are the direct targets of, and indicators of success for, cancer interventions. However, the importance of intermediate and clinical outcomes for cancer decision making rests ultimately on the extent to which they can be convincingly linked to improvement in final outcomes such as a reduction in mortality.

In order for outcomes research to achieve its potential to improve cancer care delivery, three prerequisites apply: 1) technically sound and decision-relevant final outcome measures; 2) persuasive evidence about the effect of interventions on those outcomes, with due attention to the causal linkages among intermediate, clinical, and final outcomes; and 3) the willingness and ability to translate findings into information that decision makers find understandable and compelling.

This Monograph has employed a tripartite framework for categorizing and characterizing the arenas of application for cancer outcome measures; see Table 1 here and its more expansive counterpart in the paper that introduces the Monograph (1). Macro-level studies chart population trends in cancer-related outcomes and progress against the cancer burden. Meso-level studies investigate the impact of cancer and cancer-related interventions on outcomes, with a focus (depending on the study’s specific purpose) on determining the efficacy of candidate interventions under controlled circumstances or in the community; describing patterns of cancer care, including the extent to which quality-enhancing services are embraced; or identifying effective or cost-effective interventions in the context of program evaluations or priority-setting analyses. Interventions include not only specific cancer prevention, detection, or treatment services or programs but also changes in the organization, financing, or delivery of cancer care that may influence outcomes. Micro-level studies examine the use of cancer outcome measures and measurement tools to enhance the quality of information available for patient–clinician decision making.

There is no intent here to imply that the topic areas in Table 1 are to be regarded as subsets or components of some larger, all-encompassing enterprise called “cancer outcomes research.” Rather, we seek to show how the latter can contribute substantially to a variety of analyses within each of these arenas, especially regarding the measurement of outcomes that matter to decision makers.

Because this paper attempts to identify not only the major opportunities and challenges confronting cancer outcomes research but also specific pathways forward in a number of domains, it takes the form of an extended, multipart essay rather than a summary document that only highlights the tasks ahead. While the sections and subsections below proceed in a sequence that we believe is natural and cumulative, they can be pursued in any order or even selectively, depending on reader interest. That said, the remaining sections focus in turn on:

1) Macro-level analyses, examining current and potential approaches for population monitoring of trends in HRQOL, patient satisfaction with care, and economic burden.
2) Meso-level analyses, which encompass a diverse range of outcomes-oriented studies including (as subsection topics taken up in turn): intervention efficacy (in randomized trials); intervention effectiveness (in real-world, observational designs); cancer impact, with an emphasis here on cancer survivorship issues; variations in cancer care utilization and the quality of cancer care; and clinical modeling, economic evaluation, and priority setting, with an emphasis on cost-effectiveness analysis and other approaches to inform decision makers.
3) Micro-level analyses, highlighting both the challenges in us-
Table 1. Arenas of application for cancer outcome measures

<table>
<thead>
<tr>
<th>Level</th>
<th>Arenas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro</td>
<td>Population surveillance of trends in cancer-related outcomes and progress against the cancer burden</td>
</tr>
<tr>
<td>Meso</td>
<td>Descriptive and analytical studies to understand the impact of cancer, patterns of service use, and effects of interventions on cancer-related outcomes</td>
</tr>
<tr>
<td></td>
<td>A. Efficacy</td>
</tr>
<tr>
<td></td>
<td>B. Effectiveness</td>
</tr>
<tr>
<td></td>
<td>C. Cancer impact</td>
</tr>
<tr>
<td></td>
<td>D. Variations in utilization or performance, through</td>
</tr>
<tr>
<td></td>
<td>1. Patterns-of-care studies</td>
</tr>
<tr>
<td></td>
<td>2. Quality-of-care monitoring</td>
</tr>
<tr>
<td>Macro</td>
<td>Clinical modeling, evaluation, and priority setting, through</td>
</tr>
<tr>
<td></td>
<td>1. Clinical decision modeling</td>
</tr>
<tr>
<td></td>
<td>2. Economic modeling, including cost-effectiveness, cost-benefit, and cost-consequence analysis</td>
</tr>
<tr>
<td>Micro</td>
<td>Patient-clinician decision making enhanced by the use of survey instruments and other tools to monitor and predict outcomes, as a function of the intervention selected and patient risk profile and behavioral characteristics</td>
</tr>
</tbody>
</table>

ing cancer outcome measures in clinician-patient decision making and a number of specific opportunities to make progress.

4) Improving the data and methods for cancer outcomes research on the macro, meso, and micro levels, by (i) capitalizing on existing psychometric methods to strengthen the scientific basis for assessing patient-reported outcomes like HRQOL, and (ii) accelerating progress in creating electronically oriented cancer information and surveillance systems that can capture, store, and link patient-level data rapidly and accurately while meeting confidentiality and privacy concerns.

5) Understanding the importance of, and the challenges in identifying the specific impacts of, cancer outcomes research on clinical practices policies, on cancer care delivery in the community, and ultimately on health outcomes. Such research would augment outcomes research on other conditions (2). A comprehensive discussion of this important, underemphasized component of the cancer outcomes research agenda is beyond the scope of this paper (and would merit its own book or monograph). But to illustrate how one might begin to trace the dynamic linkages from outcomes of research investigations to policies to practices to outcomes, we develop a brief case study involving screening mammography for breast cancer.

6) Emphasizing, in conclusion, that the enduring charge and challenge of cancer outcomes research is to facilitate the delivery of the right information, at the right time, to the right decision makers across the arenas of application. We need a broader, deeper understanding of where this is being accomplished, where it is not, and how to do better.

**Macro-level: Approaches to Expanding the Range of Cancer-related Outcomes in Population Surveillance**

A comprehensive assessment of the progress being made to reduce the burden of suffering and death due to cancer requires tracking population trends in cancer incidence, survival, and key patient-centered outcomes such as HRQOL, satisfaction with care, and economic burden; see Clauser in this Monograph (3). At the moment, we know much more about progress in the United States to reduce cancer-related mortality than the suffering felt by those living with a cancer diagnosis. In what follows, we discuss several possible approaches for expanding the range of outcomes that might be routinely tracked and analyzed in cancer surveillance, focusing on the patient-centered measures noted above. As noted, the general intent of such macro-level studies is to apprise decision makers of important trends, which then would indicate the kinds of meso-level investigations needed to develop a deeper understanding of factors influencing cancer-related outcomes and how to improve them.

**HRQOL**

Although a rich mixture of generic and cancer-specific HRQOL measures are now widely used in clinical trials and observational studies (as indicated in the following section on meso-level applications), in the United States, there has been no corresponding effort to measure the HRQOL of persons with cancer on a population-wide basis (3). This gap was recognized by NCI’s “Surveillance Implementation Group,” which urged in its 1999 report that “data on additional measures (e.g., patterns of care, quality of life) of cancer burden beyond incidence, survival, and mortality...be collected within established population-based cancer registries to fully assess the Nation’s cancer burden” (4). The opportunities and challenges of accomplishing this goal have recently been noted by experts collaborating within the National Coordinating Council for Cancer Surveillance (5). As HRQOL measures have been applied more widely in macro-level studies in Europe and Canada than in the United States (3), we turn briefly now to how such efforts might be accelerated in this country. Broadly speaking, this will require making better use of existing data, collecting additional data (in a strategic fashion), or both. Because different approaches are required for HRQOL measures that attempt to incorporate population-representative preferences for alternative health outcomes versus HRQOL measures derived from non-preference-based psychometric scaling techniques (3), we discuss these two broad categories in turn.

**Non-preference-based HRQOL.** Both generic (e.g., SF-36) and cancer-specific (e.g., FACT G, EORTC QLQ C30) measures could be added as modules to the National Health Interview Survey (NHIS) and other population-based efforts. Respondents self-identifying with cancer would receive both generic and cancer-specific modules, and indeed could be administered cancer disease site–specific questions (e.g., the FACT C for persons diagnosed with colorectal cancer); see Erickson in this Monograph for a discussion of this modular approach (6). The feasibility of collecting SF-36 and SF-12 data in the general U.S. population has already been demonstrated, respectively, in the Medicare Health Outcomes Survey (MHOS), conducted by the Centers for Medicare & Medicaid Services (CMS) (7) and the Medical Expenditure Panel Survey (MEPS), sponsored by the Agency for Healthcare Research and Quality (AHRQ) (8), as well as in a number of large research studies (9). Scientists at both the American Cancer Society (10) and within NCI’s Applied Research Program are exploring the use of MHOS data to better understand the experiences of cancer patients and survivors aged 65 years and older. NCI scientists are examining the feasibility of linking MHOS and registry data from the Surveillance, Epidemiology, and End Results (SEER) Program to study the relationships among cancer stage at diagnosis, initial therapy, and subsequent HRQOL. The Institute of
Medicine’s National Cancer Policy Board has proposed an alternative, complementary strategy: Collect and analyze HRQOL data on nationally representative samples of cancer patients and survivors drawn from the rolls of population-based cancer registries (11). This approach recognizes that high-quality registries (including those supported by NCI’s SEER Program and the Centers for Disease Control and Prevention’s [CDC] National Program of Cancer Registries) are also viable sampling frames for the strategic selection of cancer patients and survivors for follow-up interviews focusing on patient-centered outcomes. Such data may be further enriched by linkage to medical records or insurance claims information.

Preference-based HRQOL. While psychometric-based HRQOL measures like the SF-36 provide useful data on the impact of disease and its treatment on a range of domains, these measures do not address the potential trade-offs between survival and quality of life. In addition, they cannot reflect possible intercultural or cross-population differences in how specific health outcomes are viewed and valued. Several promising approaches, all within reach, are available for generating population-representative estimates of preference-based measures of HRQOL for cancer patients and survivors, including health-adjusted life years (HALYs) or quality-adjusted life years (QALYs).

First, synthetic estimates of preference-based outcomes can be developed by mapping non–preference-based patient-reported outcomes from population-based surveys, like NHIS and MEPS, to corresponding preference-weighted states of health (12,13), as might be found in today’s most prominent multiattribute health utility measurement “systems”—namely, the EQ-5D, the Health Utilities Index, the HALex, and the Quality of Well-Being (QWB) Index (14–17). For example, Lawrence and Fleishman used 2000 MEPS data to predict EQ-5D scores from SF-12 scores; a simple 2-variable model accounted for about 60% of the variance in EQ-5D scores (12). Such a model can readily be used to predict EQ-5D scores for any sample of cancer patients and survivors on whom one has collected SF-12 data (and vice versa).

A second approach is to survey cancer patients and survivors in a way that facilitates direct computation of their preference-based HRQOL scores. The most common variant of this, exemplified in the MEPS, is to locate a respondent’s health state position along each dimension of a multiattribute health utility measurement system (the EQ-5D in that case), and then assign her an overall preference score based on an algorithm that uses a precomputed preference weight for each possible health state. To date, such preference weights for the major utility measurement systems noted above have been derived from samples that are either geographically limited, interviewed years ago, or non–U.S. based. Most recently, however, an AHRQ-funded study has yielded nationally representative preference weights for the EQ-5D for the U.S. population (18), which can be used for subsequent QALY calculations emerging from the MEPS and other U.S. studies using this particular preference measurement system.

A third variant on this theme is to obtain direct preference assessments from sampled cancer patients and survivors—that is, to measure the survey respondent’s own utility score rather than attaching any externally derived weight. It remains to be seen whether such a resource-intensive approach to obtaining nationally representative data is feasible.

Finally, the Disability-Adjusted Life Years (DALYs) approach, developed by the World Health Organization, offers yet another avenue for producing a type of preference-weighted summary measure of health (19), and the CDC has initiated promising work to estimate DALYs on a disease-specific basis for the U.S. population (20).

Note that all of these approaches to deriving preference-based HRQOL measures can be applied in meso-level studies, particularly (of course) in cost–utility analyses, which are discussed below.

Satisfaction With Care

The opportunities and challenges here closely parallel those that arise with non–preference-based measures of HRQOL. The most promising instrumentation for assessing patient (and survivor) perceptions of and satisfaction with health care is the AHRQ-supported Consumer Assessment of Health Plans (CAHPS) (21). It has been used successfully in a number of applications across the United States, including by CMS to evaluate patient experiences in the Medicare Managed Care program. For application to macro-level cancer studies, NCI, AHRQ, and CMS are currently exploring two particular questions. First, is the current generic version of the CAHPS adequate for use by cancer patients, or does it need to be augmented with additional cancer-specific items or reoriented generally to the cancer patient and survivor? (Clearly, the same question arises with most any disease-specific application of the instrument.) Second, can CAHPS and high-quality registry data (like those from SEER) be linked to support population-based studies in which patient satisfaction is analyzed with respect to cancer type, stage of diagnosis, initial treatment, and other variables available from both data sources?

Economic Burden

Unlike the patient-reported outcomes discussed above, there are already national-level estimates of the annual economic burden of cancer in the United States, both in terms of direct medical costs only (22) and total costs (direct medical plus the “indirect” costs reflecting productivity loss) (23). However, a number of empirical and methodological challenges remain in generating high-quality, population-based estimates of the cost of cancer for macro-level analyses and reports. In response, we need to:

1) Build on what has been learned from SEER–Medicare cost analyses to accelerate efforts to link registry data (SEER and other) with insurance claims data from other public (e.g., Medicaid) and private sources (e.g., health insurers, managed care organizations) to derive direct medical costs for the total population of interest, not only the elderly. This will not be trivial, given the presence of multiple payers within most medical market areas and their likely concerns about data sharing and data privacy. Such concerns have been heightened by the confidentiality requirements for electronic data transmission imposed by the Health Insurance Portability and Accountability Act (HIPAA) of 1996 (24).

2) Devise defensible, feasible strategies for estimating or imputing these omitted cost elements (25). Direct cost estimates derived from insurance claims data will necessarily omit goods and services not covered. For example, Medicare historically has not paid for outpatient drug and most nursing
home costs generated by enrollees [though when the Medicare Modernization Act of 2003 (26) is fully implemented in 2006, tracking outpatient drug costs should become feasible]. Insurance plans generally do not reimburse for patient “time costs,” caregiver contributions, or goods and services outside the traditional medical care system.

3) Create or augment national surveys, insurance or provider administrative data, or other data platforms to gather population-based data on the indirect costs of cancer. Currently, these productivity–loss estimates continue to emerge through multistep, assumption-rich imputation processes that, no matter how carefully crafted, require additional validation within samples of cancer patients and survivors. In the end, there may be no substitute for periodic, population-based surveys that attempt to obtain information directly on work loss, wage loss, and the impact of cancer on income and wealth. As one example, the MEPS does yield data allowing inferences about illness-related wage income loss; additional survey items would permit a more comprehensive assessment of the total impact of cancer (and, indeed, other diseases).

4) Scrutinize and refine the traditional cost-of-illness (COI) methodology used to derive estimates of the total economic burden of disease. At present, both direct medical costs and the morbidity–cost component of indirect costs are estimated through prevalence-based approaches, while the mortality–cost component of indirect costs is based on an incidence-based approach (27). A potentially more satisfactory strategy is to estimate all components from an incidence-based perspective and to then derive annual, prevalence-based costs analytically from the appropriate multiperiod model.

5) Analyze the strengths and limitations of a quite different approach to determining the economic burden of cancer nationally—one based on either revealed-preference (market based) or stated-preference estimates of the dollar value of a human life (adjusted for HRQOL deficits) to derive the total value loss attributable to the disease (28). Such “willingness to pay” approaches can yield radically different estimates of economic burden than those by COI models. We need a deeper understanding of the similarities, differences, and possible inconsistencies between these two paradigms.

Meso-Level: Enhancing the Information Base for Decision Making Across the Cancer Continuum

In what follows, we consider, in turn, the categories of descriptive and analytical studies defined in Table 1. From an outcomes research perspective, the intended aim within and across these meso-level categories is to generate findings and recommendations that could support decision making about cancer care delivery, coverage and reimbursement, regulation, standards-setting, and additional research to support our efforts to ascend the “outcomes research pyramid”(2).

Efficacy

Turning first to experimental studies to assess the impact of cancer interventions on outcomes that matter to decision makers, we focus on the challenges surrounding the selection of patient-centered outcome measures, especially of HRQOL, for use in randomized clinical trials. (Issues surrounding patient perceptions of care and economic burden, less centrally relevant to efficacy assessment, will be discussed later in this section.)

Few topics in cancer outcomes research have generated more discussion, and controversy, over the last decade than HRQOL assessment in randomized trials. This is not simply because the methodological issues are thorny and intriguing (though they are). It is also because cancer clinical investigators, whether supported by the pharmaceutical industry or government agencies such as NCI, are continually facing trial design decisions concerning the inclusion of HRQOL measures—but without explicit guidance from most trial sponsors, government regulators, or potential purchasers of cancer care products. Although such guidance may soon be forthcoming, specifically from the U.S. Food and Drug Administration (personal communication with Laurie B. Burke, U.S. FDA, June 5, 2004), the debate over whether and how to include HRQOL measures in cancer trials will likely continue for some time. Indeed, the issues and challenges in this area were major topics of discussion for the five disease-specific papers in this Monograph—Mandelblatt et al. (29), Provenzale and Gray (30), Earle (31), Collins et al. (32), and Pickard et al. (33)—as well as the cross-cutting analyses by Gotay (34) and Erickson (6). They inspired at least one recent symposium that generated a compendium of relevant papers (35) and constituted a major motivation for NCI’s decision in 2001 to create the Cancer Outcomes Measurement Working Group (COMWG) to investigate the assessment of patient-reported outcomes in cancer (36).

Accordingly, the following recurring issues require ongoing investigation if we are to gain a clearer understanding and possible consensus regarding the appropriate use of HRQOL measures in cancer trials. Virtually all these issues are pertinent as well to the kindred questions about when and how to measure HRQOL in observational studies (in which patients are not randomized to interventions but selected for inclusion based on other criteria, e.g., disease diagnosis or demographic characteristics).

Conceptual foundations. As emphasized by Patrick and Chiang (37) and Ferrans (38), there are multiple competing frameworks and models for defining HRQOL and its relationship to factors both internal and external to the individual. Although the absence of consensus at the conceptual level may not be surprising, it may nonetheless slow progress in developing consensus about how to define and develop HRQOL measures for specific purposes (e.g., the appropriate dimensions for a multidimensional assessment of HRQOL in a prostate cancer treatment trial).

Value added. HRQOL data may be said to bring “value added” when they reveal information beyond that available from traditional biomedical outcomes (e.g., survival, coded symptoms, toxicity measures) in ways that influence the overall interpretation of the study and thus may appropriately inform decision making (39). To date, the most common approach for assessing this is to re-examine the published findings of studies that have included both HRQOL and biomedical endpoints and investigate the covariation (or lack thereof) between these two broad categories of outcomes. A prominent case in point is Goodwin et al. (40), which in fact built directly upon the authors’ COMWG contribution. Future work should also seek to develop innovative research designs, including prospective analyses, to examine the separate and possibly interactive influences of HRQOL and biomedical outcomes on cancer care.
choices being made by decision makers (regulators, payers, providers, patients). This goal will require a mix of quantitative and qualitative research approaches (39).

Finally, we need to better understand the relationship between preference-based and non-preference-based measures of HRQOL in terms of the similarities and the differences in the information they convey in meso-level studies. Likewise, we need to know the comparative value added of these different approaches to HRQOL assessment for informing patient and provider decision making.

**Minimal important difference.** Important recent progress has been made in defining and interpreting what is sometimes called a “clinically meaningful difference” in an HRQOL score, and which (from a decisional standpoint) could be termed more broadly a “minimal important difference” (MID) (35). Indeed, Osoba’s survey of the field leads him to conclude that a “small perceptible meaningful” change in an HRQOL score is roughly equal to 7% of the breadth of the instrument measurement scale, bracketed perhaps by 5% and 10% (41). For enhancing understanding of what any given increment or decrement in a latent variable measure like HRQOL means, we believe there is much merit in “anchor based” analyses, which examine the statistical (correlational or predictive) relationship between the HRQOL change and associated changes in other variable(s) that may have a comparatively clearer external meaning, e.g., hours able to work on the job. Again, such analyses should be given an additional, decision-analytic orientation. For example, consider a new intervention X that conveys equivalent survival benefits as current intervention Y, but has a different quality-of-life profile than Y (e.g., greater reduction in disease symptoms but also more toxicity). What is the minimum HRQOL change score that would be associated with a patient choosing X over Y, or else Y over X (all else being equal)?

**Longitudinal measurement and interpretation.** A general issue here is what has been termed “longitudinal construct validity”(42). This term refers to the challenge of evaluating the responsiveness of HRQOL over time to an intervention by examining the change-score relationships between the HRQOL measure and other variables that are posited (based on an appropriate conceptual model) to correlate with HRQOL. The task may be complicated, if not confounded, by a phenomenon called “response shift” (43). This term connotes the degree to which the very meaning of the patient’s self-evaluation of a construct like HRQOL alters over time in response to a change in 1) the patient’s internal standards of measurement; 2) the values the patient places on the domains, or dimensions, of HRQOL; or 3) the patient’s own fundamental conception of HRQOL. Deepening our understanding of when, how, and why response shift occurs as well as the implications for measuring and interpreting HRQOL measures over time (both in trials and observational studies) is a major methodological challenge.

Moreover, we posit that the seemingly distinct HRQOL issues of value added, the MID, construct validity, and response shift are not entirely distinct. For example, for a response shift to be perceptible and meaningful, it must be greater than or equal to the MID; to understand the value added of HRQOL, we must know whether the respondent’s concept of “value” itself is shifting over time. An intriguing challenge is to develop experimental or observational designs that would allow these joint-and-separate effects on choice making to be analyzed and teased out to the extent feasible.

**Sensitivity versus comparability in HRQOL assessment.** A recurring theme in the four cancer-specific papers in this Monograph is the multiplicity of alternative (and seemingly competing) measures of HRQOL employed not only across studies but even within the same study. We believe this reflects, first of all, the absence of a consensus in the cancer outcomes measurement community about what is the scientifically most appropriate HRQOL instrument for use within a given type of study (e.g., prostate cancer treatment). Indeed, a principal reason for establishing the Cancer Outcomes Measurement Working Group was to gain a deeper understanding of the psychometric strengths and limitations of particular HRQOL measures in specific types of applications; clearly, a diversity of findings and viewpoints emerged (39).

Beyond the ongoing debate about technical performance characteristics, however, is a second factor that may account for the diversity in HRQOL instrument selection: the understandable penchant for individual investigators to select measures they believe have the capacity to show substantial intervention effects in their studies. That HRQOL measure(s) must be “relevant to the study population” is not an uncommon observation from the literature or the conference podium. Yet, to the extent individual investigators elect to optimize the sensitivity of HRQOL measurement in their own studies—and in doing so, end up selecting a wide variety of measures—our ability to conduct rigorous, quantitative comparisons of findings across studies will be compromised. That is why when the NCI launched its quality-of-cancer-care initiative in 1999, it proposed to support development of a “core” set of outcome measures (44).

At this point, there appear to be at least two promising approaches for addressing the evident tradeoffs between sensitivity and comparability. First, an important feature of Erickson’s “health outcomes framework” (6) is the proposal for core measures of HRQOL that, for each particular application, may be supplemented by a module of additional items especially tailored to that cancer population. In fact, this modular approach to HRQOL assessment has already been embraced by several investigator teams [e.g., those led by Neil Aaronson (45), David Cella (46), and Charles Cleeland (47)]. For example, for evaluating HRQOL in colorectal cancer patients, the (general cancer instrument) FACT G is supplemented with items to create the FACT C. The second broad approach for dealing with the sensitivity versus comparability tradeoff is to change the nature of the “instrument” used. That is, rather than choosing between HRQOL instruments A and B (each with a fixed number of items), an investigator (using computer-adaptive methods) would draw survey questions from an “item bank” consisting of high-quality items taken from A, B, and other instruments or sources. Studies whose HRQOL assessments are based on items from the same bank will be comparable and can be appropriately sensitive, if the item bank is well constructed. How such an item bank might be created, based on application of modern psychometric methods like item response theory modeling, will be discussed in a later section.

**Effectiveness**

The focus here is outcomes assessment in observational investigations of interventions to diagnose, treat, and prevent cancer. Specifically, the question of interest is the impact of interventions on final outcomes that matter to patients, families, and other decision makers within the diverse array of care delivery
settings constituting what is often termed “community practice.” (Observational studies may also evaluate the burden of cancer or examine patterns of care, including the quality of care, delivered in the community, as discussed in the next two subsections.)

As with efficacy studies, a frequently important patient-centered outcome in effectiveness studies is HRQOL. In addition, however, measures of patient perceptions of and satisfaction with their care are being increasingly used in studies of the effectiveness of community-based interventions and health care delivery systems. (We typically do not find this form of patient-centered outcome measure in clinical trials, where the modes, methods, and even locations of care delivery are designed to conform to experimental design and are not intended to mirror experiences in community practice.) One other notable dimension of effectiveness studies is the array of statistical challenges arising from their nonrandomized study designs and other real-world factors that would be controlled for in a well-designed trial. We briefly take up these matters in turn.

**HRQOL.** Virtually all issues and points noted above under “Efficacy” are equally relevant to studies examining the effectiveness of interventions in real-world cancer care delivery settings. The challenges discussed under “Longitudinal Measurement and Interpretation” are, if anything, a greater concern in observational studies than clinical trials. Many of the latter are time-limited (e.g., 6 months, 1 year), whereas in the former, it is not unusual for data collection to continue some years beyond the index diagnosis of cancer.

**Patient perceptions of and satisfaction with cancer care.** Opportunities abound for advancing the state of the science in this comparatively new but growing area of outcomes assessment (48). Additional work is needed on:

- **Conceptual issues.** At present, there is not a reigning conceptual paradigm supporting measurement of perceptions/satisfaction, and there has been considerably less discussion about this than with HRQOL (49). Future research should investigate development of 1) a core set of domains, and accompanying core set of survey items, that cut across the cancer continuum (prevention, screening, detection, diagnosis and treatment, survivorship, and end of life) and 2) additional subdomains, and accompanying survey items, applicable to each phase of care. This would be broadly similar to the modular approach for assessing HRQOL just discussed above.

- **Methodological issues.** A number of important, practical questions require exploration. Over what timeframe can cancer patients provide valid and reliable information about their care experiences (e.g., last visit, last 6 months, last year)? Should the respondent be asked to report on or rate each component of care (e.g., provider) separately, or is it sufficient, or perhaps better, to seek more holistic assessments of the cancer care experience that “average” across many events? How does one detect and eliminate ceiling and floor effects in rating the experience of care (clearly a major issues for HRQOL, as well)? A coordinated approach to these and other issues should build squarely on the ongoing work being carried out by AHRQ scientists to refine and extend the CAHPS family of survey instruments. Cognitive testing techniques should be employed in validating candidate survey items for each posited domain, various response formats, and the impact of alternative frames of reference. [And, again, this is equally true whenever one is testing and improving the content validity of patient-reported outcomes, including of course HRQOL (50).]

**Application issues.** As the work described above proceeds, pilot tests should be conducted with cancer patients, survivors, and those at risk for cancer to evaluate the psychometric properties of these emerging new or modified measures. One prototype for such efforts is NCI’s Assessment of Patients’ Experience of Cancer Care (APECC) study currently testing CAHPS-type survey items in a cancer survivor population (51). Once core and phase-specific measures have been pilot tested, they should be applied in large-scale studies to evaluate the outcomes and quality of cancer care. The feasibility of doing this is now being examined on a limited basis by patient surveys conducted by the Cancer Care Outcomes Research and Surveillance Consortium (CanCORS), supported by the NCI and the U.S. Department of Veterans Affairs (52).

**Statistical challenges.** As noted, one of the posited prerequisites for outcomes research to achieve its potential is the capability to provide convincing evidence about the impact of interventions on the outcomes of interest. To be sure, there are important statistical design and analysis issues in clinical trials examining efficacy, e.g., loss to follow-up, which may be informative. But these challenges are magnified in observational studies, where subjects are not randomized to interventions and investigators frequently must collect data in uncontrolled circumstances (subject to myriad real-world constraints) or, in some cases, work with data previously collected for purposes other than the study at hand.

The threats to inferential and predictive validity arising from these factors include the following: 1) sample selection biases potentially affecting the distribution of both subjects and providers across interventions; 2) gaps in the information available about individual subjects reflecting either late entry into the data base (left censoring) or loss to follow-up (right censoring), both of which may be “informative” and therefore potentially biasing estimated intervention effects; and 3) the clustering or nesting of individuals within clinical teams, institutions, and geographic units. Such nesting means that observations have in fact a certain degree of connectedness, with implications for the nature and amount of unique information each data point conveys.

It is beyond the scope of this paper to discuss these matters in depth; but we strongly encourage cancer outcomes researchers to accelerate efforts already underway to explore innovative statistical design and modeling approaches to address these challenges. Such approaches might include the use of propensity scoring and instrumental variables to attempt to correct for selection biases (53); the application of parametric and semiparametric “survival” models in response to patient censoring generally (and not just due to death) (54); and the use of hierarchical regression analysis (including fixed- and random-effects “mixed” models) to tease out patient-, provider-, and geographic-level effects on outcomes in the face of nested data (55).

**Cancer Impact—Focusing on Survivorship**

As Table 1 in the Monograph’s introductory paper indicates (1), an important category of meso-level studies deals with measuring, understanding, and ameliorating the burden of cancer for patients, survivors, and those facing end-of-life decisions. In particular, there has been growing interest in better comprehending and improving the quality of life of cancer survivors, with a heightened focus on the possible late effects of disease and interventions. We examine now some of the potential contributions of outcomes research to this topic area.

Journal of the National Cancer Institute Monographs No. 33, 2004
Progress in three important areas would greatly advance both our understanding of as well as our ability to monitor and improve outcomes for cancer survivors.

**Improved tracking of the current health and cancer status of those living with a history of cancer.** Currently, the main source of information on the prevalent cancer population, currently estimated as numbering 9.8 million survivors, is derived from SEER registries (56). These registries cannot tell us, however, where individuals are in their illness trajectory, i.e., whether they are newly diagnosed and awaiting treatment, in active therapy, post-treatment and disease free, or struggling with a recurrence or advancing disease. Nor do these data provide information on the health status of those individuals. For example, we do not know among those post-treatment if they are symptom free or living with major disability as a function of their cancer. If we were able to better specify the number of people living well or poorly with their illness, we would have a means to measure the true impact of cancer not only on individuals but also on society. We would also be better able to measure the effectiveness of interventions to reduce this burden.

Because current health status (which is predictive of the clinical course of a patient’s disease) is best provided by the survivor or a knowledgeable proxy, the major barriers to accessing these needed data are twofold. First is the cost of acquiring the information (including personnel and data management systems and the resources needed to track individual survivors). Second are the restrictions imposed by current confidentiality concerns and regulations (e.g., HIPAA). Brief self-report measures of HRQOL already exist, so that it would not be difficult to locate a survivor at a specific point in an illness continuum (or ask him or her to self-identify that point). An investment in pilot efforts to capture this information in a representative sample of cancer survivors using selected tumor registries would allow inferences about the feasibility and cost of doing this on a larger scale. In this regard, statistical sampling approaches would be needed to develop new estimation models for projecting outcome distributions (for HRQOL and location on the trajectory) to the larger prevalent population (57).

Better instruments for comparing the comorbidity and disease/illness burden of populations with and without a history of cancer. Historically, most cancer survivor outcome studies have compared one group of survivors with other groups receiving different types of therapies or during different periods of illness and recovery. As cancer has become a curable or controllable illness for many, researchers increasingly are seeking to understand the impact of having cancer per se compared with another chronic and/or life-threatening condition. As this research matures, we will need studies that can tease out the relative contribution of multiple comorbidities, including those that may predate a cancer diagnosis, as well as those resulting from the cancer or its treatment. Again, this comparative information is critical if we are to understand and address the burden of cancer on survivors and society. A key question that remains largely unanswered concerns the potentially unique contribution of cancer to the health status of those living long-term with these diseases.

The principal barrier to work in this area has been the lack of instruments to adequately measure comorbidity among cancer survivors and the absence of consensus about the relative merits of available instruments. Although progress is being made in this arena, a general approach has yet to be embraced (58). A further barrier to progress is the tendency for diseases to be researched and addressed in “silos.” This issue is reflected in, and reinforced by, the very structure of the NIH with its historic orientation toward disease-specific institutes. Consensus building will be needed to drive forward efforts to capture this information. In addition, trialists should be given incentives to adopt protocols that capture comorbidity data in a routine and systematic fashion for those in the survivorship period. For example, researchers might use standard measures at set time points, such as at the start and end of treatment and perhaps one year later.

**Identifying optimal models for follow-up care.** In general, we do not have an adequate understanding of what oncology specialists and primary care physicians regard as “usual” or “appropriate” for cancer survivors. For cancer patients as a whole, 64% of those treated as adults will live 5 or more years, and 75% of treated children will live 10 years beyond diagnosis (56). Consequently, recommendations for cost-effective follow-up care is a growing concern. Yet few evidence-based guidelines exist for the care of cancer survivors, with breast cancer being a significant exception (though even there the focus is on surveillance practices alone, e.g., monitoring for cancer recurrence or second primary tumor) (59). Care needs to extend beyond mere surveillance to address problems arising from cancer-related morbidity (60). Do survivors need to return to a cancer clinic or specialist to receive this care? What tests should be done routinely? Do interventions exist or can they be developed to reduce long-term and late consequences of treatment? How can these be most efficiently delivered? Within the pediatric arena, more than two dozen specialty clinics and programs have sprung up over the last 2 decades to meet the follow-up care needs of these maturing survivors (61). Yet to date, none are evidence driven, nor do they share a common structure or document the effectiveness or cost-effectiveness of what they deliver.

Just as we lack information about survivors’ experience of care, we also lack information about survivors’ understanding of the need for follow-up care. Finally, we need to know much more about the impact of specific interventions on morbidity or mortality outcomes for survivors—that is, the effectiveness of interventions (e.g., late effects or to alleviate symptoms) in the aftermath of cancer. Until we have a better understanding of these several factors, we lack a sound starting point for the rational design of long-term care for individuals with a history of cancer or the impact of different delivery models to determine what works best to improve outcomes while controlling the burden and cost of follow-up care.

Although the research agenda for cancer survivorship outcomes is very broad, two positive activities are now under way. First, data on physician practice patterns and patient experiences with care after cancer are currently being gathered by investigators supported by the NCI and a number of other organizations. Second, the Children’s Oncology Group (COG) has published exposure-based (e.g., chemotherapy drugs, radiation doses and organs involved, surgery received) guidelines for follow up (62). COG members and others [most notably the President’s Cancer Panel (63)] have advocated that a summary of cancer treatment be given to each patient at the end of therapy. This would help survivors and their subsequent health-care providers know more about potential exposures of concern for future health (e.g., anthracycline use and later heart problems).
Monitoring Patterns of Care and the Quality of Care

In addition to yielding information about the impact of cancer on patients and survivors and the effectiveness of interventions to diagnose, treat, and prevent cancer, there is a third broad role for observational studies: to monitor access to, use of, and the quality of cancer care in community practice. Specifically, such studies can investigate population variations in the use of state-of-the-art interventions; examine gaps between evidence-based cancer care and the care delivered in the community; and analyze racial/ethnic disparities in cancer care access, service use, and outcomes. In each instance, we are interested not only in whether there is variation but why it exists. Such inquiries contribute to our understanding of whether cancer outcomes research is having its intended impact on decision making in community practice (see the “outcomes research pyramid” discussion later in this paper).

Fortunately, there is already a rich history of such studies, sponsored and/or conducted by NCI, the American College of Surgeons (ACoS), the American Cancer Society (ACS), the American College of Radiology (ACR), the American Society of Clinical Oncology (ASCO), and a number of other organizations and investigators under various sponsorships. Since 1988, the NCI has supported nearly 70 SEER patterns-of-care/quality-of-care studies, each enhanced through medical records review to investigate the patient’s care in the community following initial therapy (64). Over this same period, the SEER–Medicare database, which combines high-quality registry data with detailed information from Medicare administrative files, has yielded more than 100 publications on patterns of care, quality of care, and resource use for cancer patients aged 65 years and older (65). Since 1990, the ACoS and ACS have carried out about 100 patient care evaluation studies using their jointly supported National Cancer Data Base (NCDB). More recently, the ACoS-sponsored Commission on Cancer has established cancer-specific Disease Site Teams that employ the NCDB, augmented typically by additional medical records data, to evaluate the quality of cancer surgical care in ACoS-approved hospitals (66). The ACR has long supported such investigations (67), and ASCO has most recently stepped to the forefront with its National Initiative on Cancer Care Quality (NICCCQ), whose centerpiece effort is an observational study to evaluate the quality of care received by large samples of patients diagnosed with breast and with colorectal cancer in five U.S. cities (68).

The NCI, more than 100 SEER “special studies”—which link registry data with medical records and/or patient surveys—have been conducted over the past 15 years (69). One such study—the Prostate Cancer Outcomes Study—represents an archetype for investigating the relationship between patterns of care and patient-reported outcomes over time in community settings (70). The knowledge obtained paved the way for the NCCORS initiative to investigate patterns of care and outcomes for large cohorts of newly diagnosed lung cancer and colorectal cancer patients in diverse practice settings across the United States (52).

We catalogue these important ongoing efforts to emphasize that the stage is well set for a future generation of studies that:

1) Capitalize on high-quality registries supported by SEER, the NCDB, and the CDC’s National Program of Cancer Registries (71), as well as linked databases (such as SEER–Medicare) to enhance our knowledge of service utilization over time for an expanded portfolio of quality-of-care studies. In addition to the most prevalent cancers, it is both feasible and important to focus on a wider set of tumor types, with an eye toward guideline-endorsed interventions that may or may not now be used in community practice. In such studies, it is important not only to examine the screening, diagnosis, and initial treatment portions of the cancer continuum but also to focus attention on the survivorship and end-of-life phases, as well as on care to prevent recurrence.

2) Capitalize also on the emergence of new evidence-based measures of cancer care quality, which can feed directly into patterns-of-care studies that seek to compare community practice against recognized quality benchmarks. A major public–private project—spearheaded by the NCI and cosponsored also by AHRQ, CDC, and CMS—to identify cancer quality measures is currently being carried out by the nonprofit National Quality Forum (72). The initial focus of this project is on measures for breast cancer treatment and diagnosis, colorectal cancer treatment and diagnosis, and symptom management across the cancer continuum and end-of-life care. A major future objective of patterns-of-care studies should be to investigate the impact of such broadly endorsed “voluntary consensus standards” for cancer care quality on the delivery of cancer care, both overall and within subpopulations.

Our ability to support an expanded portfolio of such patterns-of-care studies will be enhanced to the extent we make progress toward a “national cancer data system” that links high-quality patient-level information from multiple sources (see the section below on “Developing a Cancer Information and Surveillance System”).

Clinical Modeling, Economic Evaluation, and Priority Setting

Under this rather broad heading, we address two particularly important topics: the measurement of economic cost and economic burden more generally (an important category of patient-centered outcomes in many meso-level studies) as well as the economic evaluation of cancer interventions, including the application of cost-effectiveness analysis and its cost–utility analysis variant.

Economic cost and burden. As Fryback and Craig note in this Monograph (25), there is virtually no disagreement about the conceptual underpinnings for defining and valuing the resource costs imposed by cancer and the interventions to prevent, diagnose, and treat it. Every good or service consumed because of cancer (or any other disease) should be appraised in terms of its economic opportunity cost, reflecting the value of the associated resources in their next best alternative use. But as these authors emphasize, there is wide variability across studies in how costs are defined and measured. Although there is no consensus about how to measure HRQOL in cancer care studies, there are at least high-quality instruments vying for use. By contrast, for cost assessment in cancer clinical trials or observational studies, there are no such standardized instruments. Rather, investigators today generally construct their data collection tools and algorithms based on questions from some (self-selected) set of previous studies, with new items added often in an ad hoc fashion.

To promote greater comparability across studies—as well as accuracy and comprehensiveness of cost assessment within studies—we should pursue the development of core measures of the costs of cancer and cancer care, with accompanying standard-
ized instrumentation. As with HRQOL, it is unlikely that precisely the same cost instrument will be appropriate for each and every cancer study; rather, one can envision that appropriate subsets of the total package of core measures would be applicable, depending on the nature of the cost assessment. (For example, a core measure of the cost of caregiver services would be employed in a given study if, and only if, such services were actually consumed.)

Closely related to the task of developing standardized cost measures is the delineation of (mutually exclusive) categories of resource consumption that, together, yield a satisfactory template for defining and measuring the specific aspects of the economic burden of cancer. In that regard, Hornbrook (73) has proposed a typology whose broad categories include formal medical care costs (which would support “micro-costing” analyses when resources information can be collected in detail), formal long-term care costs, lost productivity and household costs, and other societal costs.

Economic evaluations. There is large and growing literature on methodological and empirical issues in evaluating whether a given health care intervention is “worth it” from an economic standpoint (74–76). Currently, the most commonly adopted approach is cost-effectiveness analysis (CEA); if effectiveness is defined in terms of QALYs (or some other preference-based, HALY-like measure), the CEA is then generally termed a cost-utility analysis (CUA). The economic evaluation of cancer interventions can be improved in a number of ways, e.g., better estimates of economic costs, selection of QALY measures adequately sensitive to health status changes realized by cancer patients, and general adherence to good-practice guidelines for conducting such analyses (74). But we believe the following points deserve particular attention:

Most CEAs and CUA s (and not just those applied to cancer) abstract from the complex organizational and administrative realities of health care delivery and decision making. For example, such analyses generally don’t account for substantial variability in performance capability of personnel, scheduling difficulties or other variations in the process of care. They typically ignore the fact that many interventions have substantial start-up costs, which act as a substantial barrier to their adoption even when the CEA model (which usually glides over the fixed-variable cost distinction) says they are “cost-effective.” Indeed, there are a host of other real-world constraints and exigencies that influence cancer care decision making both at the planning stage and on the ground. We need research on how to enhance CEA and CUA modeling and application in ways that account for such factors.

In selecting the optimal strategy for preventing, detecting, diagnosing, or treating a chronic disease like cancer, the decision maker arguably cares about the impact of each candidate intervention on survival, HRQOL, economic burden, and other outcomes of interest from a lifetime perspective. If so, the CEA/CUA informing this decision should likewise take a lifetime perspective—even if, as often the case in cancer, the clinical trials reporting on intervention efficacy are of much shorter duration (e.g., 1–2 years, or even only a few months). To accomplish this, as both Fryback and Craig (25) and O’Brien (77) have pointed out, we need CEA/CUA models that can combine clinical trial outcomes data with information drawn strategically from observational studies and other sources to predict the lifetime flows of health and economic outcomes for each candidate intervention. For a CUA, this would allow us to compute the cost per QALY gained for each candidate, relative to the selected comparator intervention, from a lifetime perspective while drawing empirical strength from the full ensemble of available data.

An interesting recent example of using trial data to inform a CUA analysis taking a lifetime perspective is provided by Ramsay et al. (78). In this vein, O’Brien (77) argues that Bayesian statistical modeling offers a unifying approach to the problem, providing a natural platform for assessing both the uncertainty associated with each CEA/CUA estimate and the value of additional information that might, at some cost, be brought into the analysis.

An alternative approach to the economic evaluation of health care programs, which is the economist’s traditional “first approach” to such problems, is cost-benefit analysis (CBA). Long eschewed by many health care analysts and policy makers because of concerns about “placing a dollar value” on human life, CBA is steadily finding its way back into the arsenal of health economists and decision scientists; for an excellent summary of the recent literature, see Krupnick (79). In fact, CBA is being applied now in a number of disease areas, though there have been very few recent studies in cancer; see, however, Orgeta et al. (80) and Gyrd-Hansen (81) for interesting applications. The most common variant of CBA now for health program evaluation is the contingent valuation (or willingness to pay) approach: typically, decision makers are asked how much they are willing to pay to achieve a certain specified health outcome, which generally is the outcome predicted to occur, given the intervention in question. Based on the responses to such questions one derives an estimate of the total benefit of the intervention; if this exceeds the intervention’s total costs, it passes the cost–benefit test. To simplify a bit, in passing from a CUA to CBA evaluation of an intervention, such a willingness-to-pay benefit estimate replaces the QALY estimate.

Although willingness to pay may be positively related to ability to pay and thus have equity implications, there are good reasons not to dismiss this approach out of hand. First, contingent valuation techniques may have the capacity to tap into certain aspects of intervention benefit that are difficult for QALYs, as currently constructed, to register. Consequently, CBA and CEA/CUA may yield different conclusions about the merits of a candidate intervention. Second, interventions selected by a well-constructed CBA may be regarded as “economically efficient” on the basis of well-known criteria (i.e., the Pareto criterion5). This is not the case with interventions selected through CEA/CUA, unless certain additional assumptions are imposed [which, many would argue, moves the evaluation implicitly closer to CBA anyway (82–83)]. Consequently, the cancer outcomes research agenda should consider encouraging a critical exploration of both of these approaches to measuring the value of an intervention relative to its cost.

**Micro-Level: Investigating the Routine Use of Patient-Reported Outcomes in Clinical Oncology Practice**

**Challenges**

As Donaldson (84) indicates, routine assessment of HRQOL in oncology practice is rare, despite promising reports of feasibility studies. Developmental work has just begun to adapt instruments and platforms for clinical use as contrasted to their use in clinical trials. Meso-level applications of HRQOL data in-
include observational studies of HRQOL in defined populations to expand our knowledge about the effectiveness of a variety of interventions, the natural course of illness, and the quality of care. Routine use in clinical practice, however, is intended for a different purpose, that of monitoring and assisting in decision making for cancer patients. For this reason, the instruments that may be valid and reliable for research use in a clinical trial or observational study may not be when the aim is to guide and improve the outcomes of patient care rather than observe them.

For such micro-level purposes, challenges include establishing instruments’ validity for tracking individuals rather than groups, inclusion of all domains of interest to patients, clinical responsiveness and the meaning of score changes in the context of therapeutic and supportive care management, and their reliability for tracking individuals over time. All such challenges are at level 1 of the outcomes pyramid (Fig. 1). Widespread use in routine practice, like the widespread use of any technology, may be affected by research evidence that its use has value in improving patient care and patient outcomes. Such research needs to be linked to changes in policy and clinical practice before population level change can be achieved. At level 2, policy challenges could include the successful pursuit of reimbursement for the costs of collecting and using the data to address the problems identified; the development of professional guidelines and standards for their use; and the creation of strategies to protect data confidentiality.

Level 3 changes in practice, the focus of Donaldson’s paper, will be affected by many variables, including the knowledge, attitude, and behavior of clinicians; how HRQOL information is incorporated into clinic work flow; and the availability of resources for making all this happen. The factors that affect the translation of knowledge (about the value of HRQOL measurement) into clinical practice are emblematic of the many issues that influence the dissemination and use of new knowledge and technologies more generally. Ideally, the need for a better understanding of effective translation interventions will stimulate further research. In the meantime, however, experience suggests a number of opportunities to broaden and accelerate the use of HRQOL in practice.

Fig. 1. Outcomes Research Pyramid. Modified from “The outcomes of outcomes research at AHCPR” by Tunis and Stryer (2).

### Opportunities

Some opportunities lie in work that can be supported by the NCI, foundations, and other groups whether public- or private sector. Additional opportunities arise in other federal, state, private-sector, certification, and legal domains. The following deserve particular attention:

1. Syntheses of research and active dissemination of interventions that are effective in improving HRQOL;
2. Research to identify the value of HRQOL outcomes measurement in clinical practice;
3. Dissemination of clinical HRQOL measurement tools in a variety of ways that reach practising oncologists and that highlight examples of organizations that are using HRQOL;
4. Efforts to educate clinicians about the use of outcomes measurement in practice settings;
5. Translation and dissemination of work highlighting successful use of HRQOL measurement to support interventions to reduce suffering from cancer or cancer treatment;
6. Provision of technical assistance to organizations that wish to implement HRQOL measurement and to evaluate its impact, addressing not only changes in patient health outcomes but also changes in processes of care. Such research might, in turn, require the development of evaluation methods that rely less on randomized trials than quasi-experimental research designs adapted to translational research;
7. Development of new ways to provide tailored information to patients for their use in decision making. Such information for patients might include graphical formats tracking their own responses over time and providing specific findings from epidemiological studies for comparison purposes.
8. Development of platform-independent (e.g., telephone, Internet, handholds) collection and reporting applications, recognizing the varieties of ways patients and clinicians may access information, as well as differences in health literacy and preferences for inputting data and viewing reports. A flexible self-report system would give patients the freedom to choose a device that best suits their schedule, limitations, and level of health literacy.

All such applications can include links to guidelines and supportive care recommendations adapted for the individual patient, local practice, and available resources. Reports for patients can be in a format that easily allows them to monitor their own progress and to indicate when they may need immediate care. De-identified patient-reported outcome and clinical data could foster a better understanding of patterns of care and treatment effectiveness, as well as better tracking of changes in special populations and across tumor sites. Further, findings could be used over time to update the data collection instruments to improve questionnaire properties and suggest research to narrow knowledge gaps.

In addition to such developmental work and the dissemination of applications, however, we need to focus on how such micro-level innovations can be incorporated into practice flow beyond the doctor–patient visit. Use of HRQOL measurement in clinical practice can be advanced by an expanded view of the “unit of care” beyond the patient visit. A key objective is to
uncouple outcomes measurement from the strictures of the patient visit as the unit of care—an opportunity provided by information technologies that allow patients to enter and access information, when they prefer, and for clinicians to respond in a variety of formats.

To address some of these challenges, the NCI has now issued a Small Business Innovation Research (SBIR) program announcement: “Integrating Patient-Reported Outcomes in Clinical Oncology Practice”(85). The aim is to encourage development of integrated, ongoing patient-reported outcome measurement and reporting applications.

**Improve the Data and Methods for Cancer Outcomes Research**

Among the wide range of topic areas falling under this heading, we concentrate on two that have recently captured the attention of many cancer outcome researchers, and for good reason. First is the potential offered by modern psychometric approaches, such as item response theory modeling, for improving the scientific quality and feasibility of patient-reported outcome (PRO) assessment. Second is the opportunity to create high-quality, linked databases and systems that would strengthen our national capacity to monitor progress against the cancer burden and to support a wide range of meso-level observational studies on the determinants of cancer care utilization and outcomes, with a focus on disparities. Advances in both of these areas enhance cancer outcomes research at the macro, meso, and micro levels.

**Potential Contributions of Modern Psychometrics**

In our earlier discussions of future research to enhance the assessment of patient-reported outcomes (PROs), we noted the potential contributions of modern psychometrics. Now we discuss three particular opportunities for advancing the field.

**Improve cancer outcomes measurement by combining quantitative, cognitive, and qualitative approaches to measure development and evaluation.** Many HRQOL instruments (including prominent ones) have been criticized as cumbersome for respondents, not applicable over the cancer continuum or in a variety of research settings, suffering from floor and ceiling effects, and lacking a common scoring metric to allow comparisons across different HRQOL instruments. A number of survey development and evaluation tools to address these problems are available from the fields of qualitative research, cognitive aspects of survey methodology (CASM), and psychometrics. But in research to improve instrument performance they remain under-used, both singly and in combination. Integrating these methods into the design process will result in quality instruments that can be used for a number of research applications.

Qualitative researchers use focus groups to better understand the domains and issues affecting cancer patients and health-care providers (86). CASM researchers use cognitive interviewing (or cognitive testing) to assess the cognitive factors that may influence the quality of responses obtained through self-report. This involves a series of in-depth, one-on-one interviews with a small set of patients to review the organization and content of questionnaires including question comprehension and retrieval of information (50). Psychometricians use both traditional and modern measurement methods to evaluate the properties of questionnaires, and they use the information gained from both approaches to improve or develop new measures. Modern measurement theory includes item response theory (IRT) modeling, which provides a framework for analyzing item and scale properties and creating a common metric to link multiple HRQOL instruments on the same scale for combining or comparing scores (87). IRT models the relationship between a person’s health status and their likelihood of responding to each question in a scale. This item level information can be used to adapt instruments for a cancer population (88). In 2004, the NCI and the nonprofit Drug Information Association sponsored an international conference attended by outcomes researchers from government, industry, and academia to discuss the potential benefits of IRT modeling for health outcomes assessment. Later, the NCI created a website, http://outcomes.cancer.gov/conference/irt (last accessed: September 9, 2004), which includes conference presentations and findings regarding the strengths and limitations of these methods.

A key barrier to greater use of IRT modeling is the challenge of correctly understanding and applying the highly specialized software and interpreting the often arcane output. In response, the NCI has recently issued a program announcement through its SBIR funding mechanism. The SBIR announcement is intended to stimulate development of new or adapted, user-friendly IRT software that will appeal to outcomes researchers generally while being flexible enough to analyze complex data sets (88a).

**Improve validity of questionnaires translated into different languages.** Researchers have traditionally been attentive to ensuring the linguistic equivalence of PRO instruments that are translated between languages, using such methods as forward and backward translation to test for missteps. Still, translated instruments are often at risk for another, more subtle problem: populations may give culturally different responses to the same set of questions. For example, in a depression questionnaire, Azocar et al. (89) inferred that a Latino population endorsed “I feel like crying” more often than an Anglo population, arguably because Latinos regard crying as more socially acceptable behavior. This item resulted in Latinos receiving a higher average depression score than Anglos. When one group consistently responds differently to an item than another group after controlling for group mean differences on the measured construct, we have evidence of “differential item functioning” (DIF). Failure to identify scales containing such items poses a threat to the validity of between-group comparisons because scores are influenced by a variety of attributes other than those the item is intended to measure.

Concerns about DIF are not limited to translated instruments but extend to any questionnaire administered in new populations or settings other than those for which the instrument’s psychometric properties were originally evaluated. In particular, testing for and removing DIF from PRO instruments that are administered to multiple racial/ethnic and cultural groups represent potentially important contributions of psychometric science to health disparities research.

At present, few PRO analyses carry out a DIF assessment. However, interest in the topic is accelerating, as evidenced in the peer-review literature (90) and at the NCI/DIA–sponsored conference just noted above. The focus increasingly is on alternative methodological approaches based on contingency tables, IRT modeling, and structural equation analysis to assess and remove DIF in PRO measures.
Develop item banks for tailored assessment of patient-reported outcomes. Item banking represents a revolutionary new way to collect patient self-reported outcomes data. And, if executed well, it can render moot the often contentious question facing cancer outcomes researchers today: for assessing HRQOL (or any PRO), which among the several available (fixed item) instruments does one choose? The process of item banking involves two major steps. First is the pooling of items from existing instruments and the possible addition of new items (written by subject area experts) to fill clinically relevant gaps in the scale measuring any given HRQOL domain. The second step is the calibration of the resulting item pool using IRT modeling. In this way, the psychometric properties of each item in a scale, and the relationship between a cancer patient’s HRQOL level and their responses to each item, are known within the item bank.

Such banks can provide the foundation for an improved PRO assessment system for generating either short fixed forms for clinical research applications or computer-adaptive tests (CAT) that tailor the sets of items comprising the “questionnaire” administered to each respondent. CATs provide efficient and precise health status measurement by choosing, for each respondent, the most informative set of questions by selecting each question, in turn, based on the pattern of responses to all previous questions. Data collected from any short form or CAT that uses items from the same bank can be compared or combined, even when (as is typically the case) the respondents receive different sets of questions.

Although item banking and CAT-based assessment have proven to be valuable in educational assessment, there is still debate about their usefulness in health outcomes research. Analyses presented at the 2004 NCI/DIA-sponsored conference on health outcomes measurement noted above demonstrated the benefits of such tailored assessments. These benefits include reduced respondent burden (important for an aged or ill population), decreased floor and ceiling effects (important for measuring people with very poor or very good functioning), and increased sensitivity to changes in health status. In addition, several recent studies have discussed the usefulness of these methods for PRO assessment, including Cella’s report on the development of his cancer HRQOL item banks (91) and Ware’s analysis of his headache impact scale (92).

However, item banking and CAT development are both time-consuming and expensive, requiring a multidisciplinary team of experts that may in fact be difficult to assemble within one organization. A high-quality item bank requires careful, ongoing management and a commitment to regular updating as new psychometric data and analytical approaches become available. Recognizing both the promise of these new approaches and the accompanying resource demands, the National Institutes of Health (NIH) has responded with a 5-year, $25 million RFA initiative to develop the Patient Reported Outcomes Measurement Information System (PROMIS), as part of its larger Roadmap Initiative emphasizing inter-disciplinary, cross-institute projects (93). PROMIS, which will be carried out by a competitively selected network of research teams, will develop public domain item banks and CATs for key health symptom and HRQOL domains affected by a variety of chronic diseases. When operational, PROMIS will 1) deliver PRO questionnaires via a number of secure delivery platforms (e.g., computer, Internet, handheld, telephone); 2) collect PRO data for research and improvements to the system; and 3) provide instant health status reports tailored to the patient, health care provider, and researcher. Making the PROMIS available free of charge to the public will increase the likelihood of the outcomes research community adopting this new method of PRO assessment and ultimately improve PRO measurement, resulting in better medical decision making.

The emergence of PROMIS raises important issues that will necessarily be confronted in the years immediately ahead. Those issues have important implications for the cancer outcomes research agenda. For example, will the resulting bank of chronic disease survey items be sufficiently oriented to cancer PRO assessment and, if not, how do we augment the bank accordingly? (The same question holds, of course, for each major chronic disease.) How do we encourage top-flight measurement and PRO experts who were not part of the teams establishing PROMIS to engage enthusiastically in methodological and empirically oriented research that may advance the state of the science in ways not pursued or anticipated by the PROMIS investigators? Should “competitive” item banks be encouraged so as to foster critical scientific inquiry; or, would they lead to duplication of efforts while complicating the journey toward a single “core” set of HRQOL measures (for a given purpose)? These are intriguing questions that should be addressed not only by outcomes researchers, but by experts in industrial organizations, economics, property rights, and even political economy.

Developing a Cancer Information and Surveillance System To Monitor, Assess, and Improve the Outcomes of Care Across the Continuum

Current status. At present, no integrated national surveillance system of cancer incidence, mortality, prevalence, and survival exists in the United States for monitoring and supporting analyses of cancer care practices and outcomes. In some respects, however, the foundations for such an integrated system are already in place because of the evolution over the past 30 years of three national-level cancer registry efforts: NCI’s SEER Program (64), the CDC’s National Program of Cancer Registries (NPCR) (71), and the National Cancer Data Base (NCDB) (66), sponsored jointly by the American College of Surgeons and the American Cancer Society. As the IOM’s National Cancer Policy Board (11) has noted, it is feasible (and desirable) to link patient-level registry data with medical records, insurance claims and other administrative information, and even patient and provider survey data to create a detailed longitudinal profile of the cancer patient. Ideally, such a profile would begin prior to diagnosis, extend through diagnosis and initial treatment into follow-up care, and continue through the survivorship period until end of life. The value of such linked data sets in supporting analyses to understand cancer care utilization, costs, and survival outcomes has been well illustrated by the SEER-Medicare database noted earlier (94).

Yet such linked databases derived from various secondary sources are still missing information on important patient-reported outcomes like HRQOL, patient experiences with care, and economic burden (and benefit). That such PROM data can be collected and incorporated into linked data sets has been demonstrated in several recent investigations, including NCI’s PCOS (70) and ASCO’s NICQC (68). The NCI/VA–supported CanCORS Consortium (52) is capturing patient-oriented outcomes for large samples of lung and colorectal cancer patients and linking these
data with registry, medical records, and administrative information, as well as survey data from the physicians caring for each sampled patient. In addition, the American Cancer Society is sponsoring a Study of Cancer Survivors (95), which is examining a variety of outcomes for large, cross-sectional samples of survivors for six common types of cancer. A potential platform for carrying out a range of outcomes research investigations using linked data is provided by the NCI-supported HMO Cancer Research Network (CRN) (96), a confederation of research organizations affiliated with eleven large health maintenance organizations.

**Opportunities for advancement.** The key to optimizing the effectiveness and efficiency of a national cancer information and surveillance system is developing the capability to capture information rapidly and accurately from the normal, day-to-day delivery of cancer care. This has been emphasized in recent reports from the Institute of Medicine (97,98) and the U.S. National Committee on Vital and Health Statistics (99), and at the recent C-Change sponsored “Summit on Cancer Surveillance and Information Systems” (100). To the extent we make progress toward achieving such a capability, we will further four important aims:

1) Reduce the time lag between diagnosis and case ascertainment by cancer registries as new cases and their stage at diagnosis are identified as quickly as pathology reports become part of the clinical record.

2) Identify trends in health care utilization and outcomes across the cancer continuum. To do so requires shortening the elapsed time between cancer-related events and the availability of accurate statistics and expanding the menu of “core” data elements to include information on use of services (prevention, screening, treatment), patient descriptors such as co-morbidities, and patient-reported information such as HRQOL.

3) Facilitate efficient, accurate, and timely studies of patterns of care, the quality of care, and the effectiveness and cost-effectiveness of particular cancer interventions in real-world practice. Specifically, the aim—and it is a challenging one—would be to capitalize on research and administrative uses of electronic health records to significantly reduce the time lag between clinical events and the availability of information to support statistical analyses relating the processes and outcomes of cancer care.

4) Enhance opportunities for real-time feedback of clinically meaningful information to clinicians and patients to improve their decision making about cancer care. Although providing population-level data to inform individual decision making has not been within the traditional purview of cancer surveillance, such data should be important features in designing new health data systems.

Achieving these four ambitious aims could generate a wealth of information for patient care, reimbursement, regulatory, standards setting, and other policy decisions. How quickly and effectively this will happen depends on the pace of adoption of Electronic Health Record (EHR) systems (101). There are several important examples of the EHR systems, including within the U.S. Veterans Health Administration, the Indian Health Service, and a growing number of private sector provider organizations, e.g., InterMountain Health Care and Kaiser Permanente.

New data systems involving multiple, possibly regionally linked databases (rather than single data repositories) should be designed so that component databases are sufficiently standardized and interoperable to enable successful aggregation of information on cancer risk factors, diagnosis, and treatment. In particular, electronic health records are critical for supporting real-time decision making.

Reaching the four aims above will require progress on a number of fronts. Specifically, we should seek to:

1) Standardize vocabulary and messaging systems, extending work that has already been done on cancer staging and expanding approved DHHS standards, such as SNOMED (medical nomenclature) and HL7 (a standard-setting organization that is converting functions [e.g., “order entry/management”] into an operationally defined metric) (99).

2) Develop an architecture that supports database interoperability to permit gathering data across many settings and organizations. Many proponents advocate open-source architecture.

3) Address technical barriers to collecting patient-reported outcomes data. Important “enabling technologies” include the use of CAT and electronic hand-held devices or Web-based approaches for patient reporting (91).

4) Further clarify the role of HIPAA’s “privacy rule” in outcomes research (102). The rule applies to a range of health care delivery and research entities that collect and transmit “protected health information” and provides a number of ways that “covered entities” (such as health plans) can use and disclose protected health information; it needs to be more widely understood and appropriately implemented.

5) Anticipate the implications for cancer risk and prognosis assessment of our increasing knowledge about host and environment factors (and their heterogeneity) and the rapid increase in technologies to measure these factors. Ideally, the rich longitudinal data trail for each diagnosed individual would include not only traditional biomedical measures, but also information on risk factor exposure, biomarkers, and other indicators that can shed light on etiology and likely outcomes.

6) Enhance the real-time availability and value of data at the point of care to support clinician and patient decision making; for example, provide (i) trend and graphical displays of data about the patient’s clinical status, (ii) ready access to current guidelines for care, and (iii) an environment for shared decision making in which tailored information and other resources are readily available.

Yet, significant challenges remain in developing and using electronic health record systems to support outcomes research and application at the macro, meso, and micro levels. Most current EHR systems are located within separate enterprises, such as a hospital or hospital system. Although many experts and reports have called for the development of longitudinal, patient-centered systems that integrate information across these organizations, information is not yet readily transferred across most settings; for example, sharing digital images across different vendors’ applications.

Challenges notwithstanding, there are reasons for optimism. For example, though not standardized, the needed technologies are already largely available. Privacy issues, while a genuine concern, may become readily more manageable over time as researchers, payers, policy makers, and patients learn to cope successfully with the requirements of information collection,
transfer, and storage. Most importantly, there is an unprece-
dented demand by patients, payers, regulators, and policy mak-
ers for timely, accurate, and relevant information about the bur-
den of cancer—and how to reduce it. The Secretary of the
Department of Health and Human Services has recently an-
nounced a strategy for rapid implementation of electronic health
records that has captured the attention of vendors, providers of
care, insurance plans, and researchers (103). And the NCI has
embarked on a promising initiative in bioinformatics that may
encompass EHR systems and some of the necessary components
outlined as crucial for cancer outcomes research (104).

Ascending the Outcomes Research Pyramid:
An Illustration

The opening paper (1) in this Monograph (Lipscomb et al.,
2004) introduced the “outcomes research pyramid” developed
by Stryer and Tunis (2) as a simplified model positing that the
various impacts and consequences of outcomes research could
be seen as involving four different levels of activity (Fig. 1). At
level 1, research stimulates the development of additional re-
search tools, methods, and instruments, and other topics in out-
comes research. Level 2 refers to policies and programs that may
be created as a result of outcomes research. Level 3 refers to
changes in what clinicians or patients do, and level 4 refers to the
effects of research on final health outcomes, both biomedical
and patient reported.

To illustrate how this model can help frame a discussion
about the impacts of outcomes research, we select a well-known
cancer intervention—screening mammography for breast can-
cer. Mammography was first studied in a randomized trial in
the early 1960s, and dissemination occurred in the mid-1980s to
mid-1990s. By the late 1990s, breast cancer mortality had
clearly declined. Behind such a simple statement, however, lies
40 years of work involving development and research, a wide
range of policy decisions, a progression from infrequent to broad
use in clinical practice, and finally, demonstrable reductions in
population-level mortality. The story of breast cancer screening
demonstrates the recursive nature of interaction among levels,
and the evolutionary nature of changes in care.

Level 1. In 1965, Egan and others reported on the positive
value of detecting and intervening surgically on a tumor that
could not yet be felt (105), leading to the now classic random-
ized trial of mammography conducted among women in a New
York HMO (106). Although initial results of that trial were
released in 1966 and 1967 (107,108), a second generation of
studies was initiated to confirm the mortality reduction finally
reported by Shapiro and his colleagues in 1971 (109). It was the
first study of cancer screening to show results in a randomized
trial; yet, before decision makers were ready to act, there needed
to be confirmation of the findings.

Several countries sponsored randomized trials to validate the
original findings that screening reduced mortality and to address
ambiguities about the age to begin screening and the interval
between screens (110–112). The idea of encouraging women to
seek care even without symptoms also needed testing. The Breast
Cancer Detection and Demonstration Project was under-
taken to learn whether women would seek mammography
screening even though they were without breast problems (113).
As the evidence showing efficacy increased, and the willingness
of women to seek care was demonstrated, a new set of research
studies sought to understand how to promote mammography
(114). Classic work examining the effects on screening behavior
of practice guidelines, clinical reminders to physicians and to
women, as well as women’s concerns about breast cancer grew
out of this era (115–120) and continues today.

As the use of screening mammography increased, the Na-
tional Cancer Institute established the Breast Cancer Surveil-
lance Consortium (121) in 1994 to enhance understanding of
breast cancer screening practices in the United States and their
relation to changes in stage at diagnosis, survival, and breast
cancer mortality.

In the outcomes research pyramid, all such activity is repre-
sented in level 1 but, as the examples show, the screening re-
search itself has been occurring throughout the 40 years of prac-
tice change.

Level 2. At level 2 (policy change), research began to affect
policies on reimbursement, certification, and clinical screening
guidelines, such as those issued by the American Cancer Society
in 1987 (122) and the U.S. government in 1992 (123). For ex-
ample, private health plans started to establish screening guide-
lines and promote their adoption. Private insurers also began
issuing guidelines in the 1980s (124). Use of screening mam-
mography became a measure of the quality of prevention, and
the National Committee for Quality Assurance (NCQA) in-
cuded it in its set of measures to certify and compare perfor-
ance among health plans (Health Plan Employer Data and
Information Set (HEDIS) version 2.0 (125).

Policy changes continued during the 1990s as various gov-
ernment decision makers began to support efforts to increase
both the quantity and quality of screening (126,127). Reimburse-
ment for mammography became a cause for many groups, and
concerted lobbying by breast cancer advocacy organizations led
some states to require insurers to cover screening (128,129).
In 1992, Medicare made the decision to cover mammography
screening for its beneficiaries (125,126,130). Concerns about ra-
diologic quality then led to passage of the Mammography Quality
Standards Act of 1992 (Public Law 102–539) to help ensure that
facilities provided high-quality images and interpretation (131).

Level 3. Level 3 of the outcomes research pyramid addresses
the impact of research on clinical practice. Despite confirmatory
research regarding the benefit of mammographic screening in
the 1980s (108), mammography was not widely used in clinical
practice until the late 1990s. Then research, policy, and other
factors such as efforts to educate women about mammography
had succeeded in raising the level of screening throughout the
United States. In 1987, less than 30 percent of women reported
being screened by mammography within 2 years; by 1995, the
proportion screened had more than doubled, and by 2000 was
reported at 70.1% (132).

Level 4. Level 4 of the pyramid focuses on outcomes. Despite
the growth in the use of screening, mortality reductions did not
appear immediately. Implementing screening in practice was
different from implementation in trials. In practice, there needs
to be adequate capacity to reach women, and as noted above,
screening did not reach high proportions of the population of the
United States until the mid-to-late 1990s. In fact, it was not until
the mid-to-late 1990s that breast cancer mortality began to de-
crease in the United States (133).

Although that decline occurred close to the time that screen-
ing had become widespread in practice, there remain questions
about how much of the decline is actually attributable to screen-
ing. In particular, while screening was increasing, there were
important changes in treatment options, including the introduction of tamoxifen, and how much of the observed mortality reduction is due to increased screening versus better treatment is a point of discussion even now (134–137). Important insights in this regard are likely to emerge soon from the NCI-sponsored Cancer Intervention and Surveillance Network (CISNET) project, whose seven breast cancer research groups are modeling the impact of screening and adjuvant therapy on breast cancer mortality trends in the United States over the 1975–2000 period (138).

So, the story of breast cancer screening research has come full circle. It began with a key study in the 1960s, more research in the 1970s and 1980s, and policy changes throughout the 1990s. Yet we do not intend to imply through this example that the evolution of mammography policy was shaped only by research—advocacy groups, for example, were clearly influential (139–140).

Nor do we suggest that screening practices changed only because of shifts in clinical policy guidelines or reimbursement rules (since provider knowledge about the procedure was deepening generally), or that improvement in breast cancer survival outcomes are solely attributable to screening and the resulting favorable trends toward early-stage diagnosis (since more effective treatments became available during this period). Moreover, other factors such as the availability of technology, access to health care, and medical malpractice litigation, can also influence breast cancer screening practice and outcomes (141, 142).

Despite clear improvements in the use and the potential impact of mammography on mortality from breast cancer in the general U.S. population, important policy and research issues continue.

First, the benefit of mammography is not universally shared—low-income populations and those without insurance are much less likely to be screened. For example, although screening rates are climbing, and late-stage at diagnosis is decreasing among the general population of screened women, those without health insurance are much less likely to receive a mammogram and are much more likely to have late-stage disease at diagnosis (131,143–145). Further, while mortality reduction is evident in the population as a whole, the overall mortality rate among African American women is rising (145).

Second, mammography finds only about 80% of incident breast cancers, and tens of thousands of women still die of breast cancer in the United States each year (145). To improve screening sensitivity, new methods of imaging that use digital displays rather than screen-film are being investigated, and entirely new approaches like magnetic resonance imaging are on the horizon (146). Improvements in these technologies will, in turn, enhance opportunities for improvements in breast cancer final outcomes, though questions about the cost-effectiveness of these new approaches, compared with screening mammography, will require ongoing study.

Third, in virtually all screening mammography analyses to date, patient mortality and survival have been the primary final outcomes; however, patient-reported outcomes like HRQOL are also important and should be re-emphasized and better integrated into the evaluation of screening mammography (and cancer interventions generally).

Finally, in this example, we should not disregard the many other research issues that continue to be debated, such as the appropriate age range for screening mammography, the value of clinical examination combined with mammography, or how the quality of the imaging might be improved. Rather, our aim has been to illustrate that the “outcomes research pyramid” can serve as a useful model for framing a systematic investigation of how research may influence clinical and public policies, which then may influence final outcomes. As the example suggests, the journey up the pyramid is unlikely to be a simple linear one, but better characterized as recursive, interactive, and dynamic. And because the challenges presented by cancer and the opportunities offered by scientific advancement will inevitably change over time, the journey may be ever forward, but it is never ending.

A NEW FRONTIER: UNDERSTANDING AND IMPROVING THE PROCESSES AND OUTCOMES OF CANCER CARE DECISION MAKING

In its annual plan and budget document submitted to the President, the NCI has put forth a research initiative to improve the quality of cancer care, “...by strengthening the scientific basis for public and private decision making in care delivery, coverage, purchasing, regulation, and standards setting” (147). In this Monograph, we have discussed how outcomes research can contribute to the knowledge base required for better cancer care decision making, focusing on three distinct arenas of application for outcomes measurement and improvement (Table 1).

Studies at the macro level examine trends in cancer-related outcomes and progress against the cancer burden at the population level. Such surveillance efforts are intended to inform policy formation and the research agenda, particularly for investigations conducted at the meso level. Included here is a diverse range of descriptive and analytical studies whose common aim is to better understand and improve cancer outcomes at the level of the individual. Because meso-level studies may either seek to identify interventions that improve outcomes, track patterns of use of outcomes-enhancing interventions, or produce decision analyses for selecting the “best” intervention (by some specified criterion), they respond directly to the research needs underscored in the NCI’s quality-of-care goal statement. Specifically, such studies can provide empirical findings and recommendations to influence decision making by patients, families, providers, and payers, as well as by agencies and organizations that render judgments about the safety, efficacy, effectiveness, and cost-effectiveness of cancer care. Studies at the micro level use outcomes measurement to promote improvement in the quality of patient-provider communications and decision making and thus, it is posited, the quality of cancer care itself.

That such outcomes research applications can influence cancer care decision making leads quickly to the questions of whether, when, and how this happens in reality. It was precisely these questions that led (what is now) the Agency for Healthcare Research and Quality to conduct its own investigation of “the outcomes of cancer outcomes research,” as evidenced in the large portfolio of health care effectiveness research the agency supported over the 1989–97 period (2). From this report emerged the Outcomes Research Pyramid and the conclusion that health outcomes research had contributed significantly to the empirical and methodological foundations of data collection, analysis, synthesis, and application to formal models for inference and prediction. However, only “a few examples” of outcomes research (at level 1 of the pyramid) had produced findings incorporated into practice policy (level 2) or clinical decision
making (level 3), or had been shown to improve health or economic outcomes (level 4).

In the screening mammography example in the previous section, we did find suggestive evidence that breast cancer mortality outcomes in the United States were improved as a consequence of changed screening and treatment practices, as likely influenced by guidelines and other clinical standards, which in turn were shaped by research findings on the effectiveness of screening. Yet, based simply on published empirical studies, it is difficult to isolate and pinpoint the impact of specific research findings on decision making at levels 2 and 3. In tracing the influence of cancer outcomes research on decision making processes and decisional outcomes, which are then intended to improve health outcomes, it is clear that the veritable journey up the pyramid is not simple and linear but potentially much more complicated. There are likely to be interactive, recursive relationships involving levels 1–3 (e.g., between clinical policies at level 2 and research at level 1). Health outcomes (level 4) may dynamically influence the outcomes research agenda, while having a direct impact on both policies and clinical practice.

This brief (even stylized) screening mammography example nonetheless makes an important point. If we are to acquire an adequate understanding of how cancer outcomes research currently influences—and could more effectively influence—clinical policies and practices, and then outcomes, we need an in-depth understanding of the processes of decision making within and between levels 1–3 of the pyramid. In particular, we need to know how decision makers who establish cancer practice policies (e.g., clinical guidelines) seek out, evaluate, synthesize, use, reject, or adapt the findings of outcomes research studies, given the multiple factors potentially influencing choice (e.g., liability issues, patient advocacy, personal experiences, and the like). In the same way, we need a clearer picture about the impact of policies on cancer care delivery decision making.

Specifically, within levels 2 and 3 of the pyramid, the following “loci” of decision making merit in-depth study to ascertain whether cancer outcomes research findings are having an impact and, if so, in what ways. Under each of these loci, we list here either the products or tools arising within each or the type of agency or organization that is relevant to the decision being made. The task for cancer outcomes research is to probe the anatomy and physiology of how decision-relevant information is produced and used in each case.

**Clinical Policy Formation**

- Guidelines (e.g., as produced by the National Comprehensive Cancer Network (148))
- Performance benchmarks (e.g., as established by a hospital system, based on current “best practice”)
- Quality and safety performance measures (e.g., as anticipated from the Cancer Quality of Care Measures Project being conducted by the National Quality Forum (72))
- Economic assessments, including cost-effectiveness analysis (e.g., as being carried out by the United Kingdom’s National Institute for Clinical Excellence (NICE) (149) and by public and private analysts in the United States in adherence to recommendations from the Panel on Cost-Effectiveness in Health and Medicine (74) or the new guidelines issued by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (76))

**Regulation and Standards Setting**

- Product approval and marketing decisions [e.g., by the U.S. FDA and the European Medicine Evaluation Agency (EMEA) (150)]
- Evaluation of health plans and provider performance [e.g., by the National Committee for Quality Assurance (151) and the Joint Commission on the Accreditation of Healthcare Organizations (152)]

**Coverage and Reimbursement Decisions by Third Parties**

- Federal payers [e.g., Medicare national-level and “intermediary” payment decisions (153,154)]
- State-level payers (e.g., Medicaid programs)
- Private insurers (e.g., Blue Cross–Blue Shield plans)

**Purchasing, Benefits Package Design, and Care Delivery Decisions by Providers**

- Public providers of care (e.g., Department of Defense, Department of Veterans Affairs, Indian Health Service)
- Private providers of care (e.g., HMOs, and other medical care organizations large and small, including solo practitioners)
- Large purchasers of care (e.g., Federal Employee Health Benefits Program and corporations such as Ford and General Electric)
- Organizations that encourage “value purchasing” of health care services (e.g., Leapfrog Group (155), the National (formerly Washington) Business Group on Health (156), and the Academy of Managed Care Pharmacy (157), which has produced a guidance document to improve the scientific quality and comparability of drug formulary submissions)
- Patient advocacy groups (158)

**Decision making by individuals, families, and clinicians,** based on the information available from all sources and shaped (and sometimes constrained) by decisions reached at the other loci above.

To develop a deeper appreciation of the decision making processes at these levels will require us to pursue descriptive studies that, for example, correlate the publication of a cancer practice guideline with subsequent changes in patterns of care. In addition, however, we should encourage well-designed longitudinal studies to better track causal linkages and qualitative research featuring detailed case studies to probe the ways in which particular outcomes research findings influence policy decision making processes and outcomes (at level 2). Similarly, we need studies to trace the impact of these clinical policies on decision making processes and outcomes in cancer care practice (level 3). Developing a sound understanding of how such decisions are made is critically important to designing a cancer outcomes research strategy that can inform and influence the decisions. These activities represent important “frontier” areas of investigation.

**REFERENCES**


National Committee on Vital and Health Statistics. Information for health: Transforming health care through information technology at national con - 2004.]


Notes

1 Editor’s note: SEER is a set of geographically defined, population-based, central cancer registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes the data available to the public for scientific research.

2 To illustrate how CEA/CUA and CBA can readily yield different perspectives, and possibly different conclusions, about a candidate intervention, consider the following simple, entirely hypothetical example. Suppose for a patient suffering from chronic, excruciating pain there is the option of a new intervention that could provide complete relief, though for just 2 weeks; after that the pain would resume.

Assume, for simplicity, that the patient regards this excruciating pain to be “as bad as death” (implying a QALY score of 0.0 per unit of time, by typical assumptions) and that relief from the pain results in “excellent health” for those 2 weeks (implying a QALY score of 1.0 per unit of time, by typical assumptions). The effectiveness of the pain intervention, from a QALY perspective, is about 0.038 (=[(1.0−0.0)×(252)] QALYs. Its cost-effectiveness, then, hinges on the total economic cost of the intervention, which includes the cost of the drug and associated hospitalization and other related expenses. There has been much debate but no sharp consensus on the range of cost–utility ratios for which a given intervention represents a “good buy” or not, from a societal perspective.

But conventional rules-of-thumb have tended to regard CUA ratios below $100,000/QALY gained as suggesting the intervention is a candidate for adoption (the smaller the ratio, the better), while ratios larger than this raise questions about the intervention’s cost-effectiveness (the higher the ratio, the poorer the support for adoption). In the example at hand, if the pain intervention’s cost happens to be $2000, its cost-utility ratio is $52 000 per QALY gained [$200 000/$0.038], indicating a reasonably good buy, while if the cost is $4000, the CUA ratio is $104 000 per QALY, which is borderline unfavorable.

From a cost–benefit analysis perspective, the crucial question (given the data) is this: How much is some relevant combination of payers (patient, family, insurers, public agencies) willing to pay for the drug? If the answer is $4000 or greater, the intervention passes the CBA test, and vice versa. It should be clear that such CBA and CEA/CUA analyses can yield the same, or different, recommendations about an intervention. This is principally because they offer two quite different approaches to valuing health benefits. It is noteworthy that some CUA analysts have shown a growing interest in obtaining viable estimates of society’s willingness to pay for a QALY. This would allow any CUA to be converted into what proponents term a “net benefit analysis” (159), which is essentially a form of CBA.

3 Any policy action (public or private) is Pareto efficient if it leads to at least one person being made better off while no one is made worse off (which in some cases might require adequate compensation to potential losers) (160). In the pain intervention example, the assumed measure, or index, of “better off” is the amount decision maker(s) are willing to pay. If this amount is greater than or equal to the resource costs of the intervention, it is thereby possible to compensate these resource owners so that none is worse off (i.e., each is assumed to be paid the opportunity costs of the resources being contributed). In this case, adopting the intervention improves economic efficiency (from a Pareto perspective). In fact, the difference between total willingness-to-pay and total costs, sometimes called “consumer surplus,” is a rough index of the net benefit of adopting the intervention.

The authors benefited significantly from comments by Robert Coyle, PhD, Director, Division of Cancer Control and Population Sciences, National Cancer Institute; and Rachel Ballard-Barbash, MD, MPH, Associate Director, Applied Research, Division of Cancer Control and Population Sciences, National Cancer Institute.