

Supplementary Appendix 2

This appendix has been provided by the author to give readers additional information about his work.

Supplement to: Porter ME. What is value in health care? *N Engl J Med* 2010;363:2477-81. DOI: 10.1056/NEJMp1011024.

This framework paper is Supplementary Appendix 2 to

Porter ME. What is value in health care? N Engl J Med 2010; 363:2477-81
(10.1056/NEJMp1011024).

MEASURING HEALTH OUTCOMES: THE OUTCOME HIERARCHY

Michael E. Porter, Ph.D.

Achieving good patient health outcomes is the fundamental purpose of health care. Measuring, reporting, and comparing outcomes is perhaps the most important step toward unlocking rapid outcome improvement and making good choices about reducing costs. Outcomes are the true measures of quality in health care. Understanding the outcomes achieved is also critical to ensuring that cost reduction is value enhancing.¹⁻³ Thus, outcome measurement is perhaps the single most powerful tool in revamping the health care system. Yet systematic and rigorous outcome measurement remains rare or nonexistent in most settings.

There are a growing number of examples of comprehensive outcome measurement that provide evidence of its feasibility and impact. At the national level, Sweden and Denmark are the clear leaders in establishing national quality registries covering many conditions.⁴ In the United States, federal legislation has mandated universal outcome measurement and reporting by all providers in organ transplantation,⁵ in vitro fertilization,⁶ and dialysis care. At the provider level, the most advanced large-scale efforts are occurring in two German hospital groups and at some U.S. providers.⁴ Examination of these efforts leads to some clear conclusions. First, in each case, outcome measurement has proven to be practical and economically feasible. Second, accepted risk adjustment has been developed and implemented. Finally, measurement initially revealed major variation in outcomes in each case, but led to striking outcome improvement and narrowing of variation across providers over time.

The feasibility and impact of comprehensive outcome measurement is no longer in doubt. However, the current state of outcome measurement leaves much to be desired. There is no consensus on what constitutes an outcome, and the distinctions among care processes, biologic indicators, and outcomes remain unclear in practice. Outcome measurement tends to focus on the immediate results of particular procedures or interventions, rather than the overall success of the full care cycle for medical conditions or primary and preventive care. Even the best efforts are often limited to one or a small number of outcomes, frequently those that are most easily tracked. Measured outcomes often fail to capture dimensions that are highly important to patients. Finally, many outcome measurement efforts are ad hoc and not comparable across providers.

This article offers an overall framework for outcome measurement to guide the development of the full set of outcomes for any medical condition. It introduces the outcome measures hierarchy as a tool for identifying the appropriate set of outcome dimensions, specific metrics, and associated risk factors. It explores the relationships among different outcome dimensions, their weighting by patients, and the relationship of outcomes to the cost of care. I examine the process by which outcomes improve over

time as well as the evolution of risk factors. Finally, the article examines the benefits and costs of standardized or monetized outcomes across medical conditions. The detailed steps involved in creating and implementing an outcome measurement system are developed further in another article.

The Unit of Outcome Measurement

Outcomes are the results of care in terms of patients' health over time. They are distinct from care processes or interventions designed to achieve the results, and from biologic indicators that are predictors of results. However, discomfort, timelines, and complications of care are outcomes, not process measures, because they relate directly to the health status of the patient.¹ Patient satisfaction with care is a process measure, not an outcome. Patient satisfaction with health is an outcome measure.

In any field, quality should be measured from the customer's perspective, not the supplier's. In health care, outcomes should be centered on the patient, not the individual units or specialties involved in care. For specialty care, outcomes should be measured for each medical condition or set of interrelated patient medical circumstances, such as asthma, diabetes, congestive heart failure, or breast cancer. A medical condition includes common complications, coexisting conditions, or co-occurring conditions. Each medical condition will have a different set of outcomes. For primary and preventive care, outcomes should be measured for defined patient populations with similar health circumstances, such as healthy adults, disabled elderly people, or adults with defined sets of chronic conditions.

Outcomes should be measured for each medical condition covering the full cycle of care, including acute care, related complications, rehabilitation, and reoccurrences. It is the overall results that matter, not the outcome of an individual intervention or specialty (too narrow), or a single visit or care episode (too short). If a surgical procedure is performed perfectly but a patient's subsequent rehabilitation fails, for example, the outcome is poor. For chronic conditions and primary and preventive care, outcomes should be measured for periods long enough to reveal the sustainability of health and the incidence of complications and need for additional care.

Generalized outcomes, such as overall hospital or departmental infection rates, mortality rates, medication errors, or surgical complications, are too broad to permit proper evaluation of a provider's care in a way that is relevant to patients. Such generalized outcomes also obscure the causal connections between specific care processes and outcomes, since results are heavily influenced by many different actors and the specific mix of medical conditions for which care is provided.

Health care's current organizational structure and information systems make it challenging to properly measure outcomes. Thus, most providers fail to do so. Providers tend to measure only what they directly control in a particular intervention and what is easily measured, rather than what matters for outcomes. Providers also measure outcomes for the interventions and treatment they bill for, rather than outcomes relevant for the patient. Outcomes are measured for departments or billing units, rather than for the full care cycle over which value is determined. Much outcome work is currently driven by medical specialty expert or consensus panels, not by multidisciplinary groups for medical conditions. Faulty organizational structure also helps explain why physicians fail to accept joint responsibility for outcomes, defending this by their lack of control

over “outside” actors involved in care (even those in the same hospital) as well as over patient compliance.

The first step in outcome measurement is to define and delineate the set of medical conditions to be examined (or the patient populations in primary care settings). Setting medical condition boundaries requires specifying the range of related diseases, coexisting conditions, and associated complications included, as well as the beginning and end of the care cycle.

For any medical condition (or patient population in primary care), defining the relevant outcomes to measure should follow several principles. First, outcomes should involve the health circumstances most relevant to patients. Second, the set of outcomes should cover both near-term and longer-term patient health, addressing a period long enough to encompass the ultimate results of care. For chronic conditions, ongoing and sustained measurement is necessary. Third, outcomes should cover the full range of services (and providers) that jointly determine the patient’s results. Finally, outcome measurement should include sufficient measurement of risk factors or initial conditions to allow risk adjustment (see below).

The Outcome Measures Hierarchy

There are always multiple dimensions of quality for any product or service, and health care is no exception. For any medical condition or patient population, multiple outcomes collectively define success. The set of outcomes is invariably broad, ranging from immediate procedural outcomes, to longer-term functional status, to recovery time, to complications and recurrences. Survival is just one outcome, albeit an important one, as is the incidence of particular complications or medical errors. Medicine’s complexity means that competing outcomes (e.g., near-term safety and long-term functionality) must often be weighed against each other.

The full set of outcomes for any medical condition can be arrayed in a three-tiered hierarchy (see Figure 1). The top tier of outcomes is generally the most important, with lower-tier outcomes reflecting a progression of results contingent on success at higher tiers.

Each tier of the hierarchy contains two broad levels, each of which involves one or more distinct outcome *dimensions*. Outcome dimensions capture specific aspects of patient health. These outcome dimensions are the critical dimensions of quality in health care. For each dimension, success is measured with one or more specific *measures* or *metrics*. Finally, for each measure there are often several choices in terms of the *timing* and *frequency* of when to measure it.

Tier 1 of the hierarchy is *patient health status achieved*, or for patients with some degenerative conditions, health status retained. The first level, survival, is of overriding importance to most patients. Survival (or mortality) can be measured over a range of periods appropriate to the medical condition. For cancer, 1-year and 5-year survival are common metrics. Maximizing the duration of survival may not always be the most important outcome, however, especially for older patients who may weight other outcomes more heavily. I discuss the weighting of outcomes below.

Effective outcome-measurement systems must move well beyond survival, because survival alone omits many factors of great significance to patients. (Note that survival is sometimes used as a proxy for the broader effectiveness of care.) Measuring

the full set of outcomes is also essential in order to reveal the connections between care processes or pathways and patient results.

The second level in Tier 1 is the *degree of health or recovery* achieved or retained. Regaining or preserving health is the ultimate purpose of most health care, with the exception of end-of-life or palliative services. Level two should capture the *peak* or *best steady-state* level of health achieved, defined according to the condition. Degree of health or recovery normally includes multiple dimensions such as freedom from disease and relevant aspects of functional status. For head and neck cancer, for example, level two outcomes include not only whether remission is achieved, but functional outcomes such as the ability to eat and speak normally, maintain appearance, and avoid depression.⁷

Tier 2 of the outcomes hierarchy is the *process of recovery*. Recovery, or the process of achieving the best steady-state level of health attainable, can be protracted and arduous. Reducing the duration, complexity, and discomfort of recovery, in a manner consistent with achieving good Tier 1 outcomes, constitutes another group of important patient results.

The first level in Tier 2 is the *time* required to achieve recovery and return to normal or best attainable function. This can be divided into the time needed to complete various phases of care, such as time to diagnosis, time to treatment plan, time to care initiation, and duration of treatment. Cycle time is an outcome with major importance to patients, not a secondary process measure. Reducing cycle time yields direct benefits to the patient in terms of reducing the burden of recovery and can also affect health status achieved and its sustainability. For example, rapid initiation of therapy and avoidance of interruptions in therapy are often major influencers of prognosis in patients with cancer; after a myocardial infarction, faster time to reperfusion can improve function and reduce complications. The relationship between cycle time and health status achieved is just one of many instances in which outcomes at one level in the hierarchy can affect outcomes at other levels (see below).

The second level in Tier 2 is the *disutility of the care process* in terms of missed diagnosis, failed treatment, anxiety, discomfort, ability to work or function normally while undergoing treatment, short-term complications, retreatment, and errors, together with their consequences. This level can cover a wide range of dimensions depending on the condition. Ineffective or inappropriate treatments that fail to improve health will show up here, as will medical errors and treatment complications that lead to interruptions in care. Disutility of care will frequently affect the timeline of care.

Tier 3 is the *sustainability of health*. Sustainability measures the degree of health maintained as well as the extent and timing of related recurrences and consequences. The first level in Tier 3 is recurrences of the original disease or associated longer-term complications. Measures of time to recurrence and the seriousness of recurrence would fall here. The second level in Tier 3 captures *new* health problems created as a *consequence of the treatment itself*, or care-induced illnesses. When recurrences or new illnesses occur, some higher-tier outcome dimensions such as survival, degree of recovery from the recurrence, and so on, will also apply to measuring the outcome of these recurrences or illnesses (see the dotted lines in Figure 1).

With some conditions, such as metastatic cancers, providers may have limited impact on survival or other Tier 1 outcomes, or survival rates may be uniformly high. In

these cases, providers can differentiate themselves on Tiers 2 and 3 by making care more timely, reducing discomfort, or limiting recurrences.

Defining Specific Outcome Dimensions and Measures

Each medical condition (or population of primary care patients) will have its own unique set of outcome measures. The importance of each tier, level, and dimension of outcomes will vary according to medical condition and sometimes according to the subgroup of patients. For most conditions, there will be multiple outcome dimensions at each level (with the possible exception of care-induced illness). The number of dimensions at each level will depend on the range of complications, the variety of treatment options, the duration of care, and so on. Broadly defined outcome concepts, such as functional status, must be subdivided into specific dimensions that are relevant to the condition. For example, rather than apply a generic activities of daily living assessment to all patients upon hospital discharge, the ability to eat and speak normally could be added to the measures tracked following head and neck cancer treatment.

Each outcome dimension may involve one or more specific measures and multiple periods. Survival is a single dimension, for example, but can be measured in a variety of ways and for several relevant periods. These choices will depend on the medical condition or patient population.

Selecting Outcome Dimensions

Figure 2 provides illustrative sets of outcome dimensions for breast cancer and acute knee osteoarthritis requiring replacement. These examples are not meant to be exhaustive, but to illustrate the structure for the comprehensive sets of outcome dimensions that are needed to fully describe patients' results — which most current measurement efforts fail to capture. No known organization systematically measures the entire outcome hierarchy for the medical conditions it addresses, though some are making good progress.

There are inevitably choices involved in selecting the set of outcome dimensions to measure. The most important criteria in making these choices should be importance to the patient, variability, frequency, and practicality. The outcome dimensions chosen should be important to the patient. Engaging patients and their families in defining this importance is an invaluable step, through focus groups, patient advisory councils, or other means. Outcome dimensions should be variable enough to require focus and improvement. Thus adverse outcomes chosen for measurement should occur often enough to justify the costs of measurement, though very rare outcomes must be measured if they are very important to the patient. The practicality of accurate measurement must also play a role in determining what to measure, as noted above. Controllability, or the provider's current ability to affect the outcome, should be secondary because the key purpose of outcome measurement is to document problems that need to be studied and addressed.

At their outset, outcome-measurement efforts should include at least one outcome dimension at each tier of the hierarchy, and ideally one at each level. As experience and data infrastructure grow, the number of dimensions (and measures) can be expanded over time.

Relating Outcomes to Processes

To identify the set of outcome dimensions, a useful approach is to chart the cycle of care for the medical condition being examined. The care delivery value chain (CDVC), shown in Figure 3 for breast cancer, is a tool for mapping the full set of activities or processes involved in care.⁸ The CDVC highlights the full care cycle and all the involved entities or units. This full map of the care cycle allows a systematic identification of all the relevant outcome dimensions as well as when and where measurement should occur.

The CDVC not only helps to identify dimensions and measures, but also enables particular outcome dimensions to be linked to the specific processes of care from which they arise. The connections between the CDVC and outcomes, then, are important to guiding outcome improvement.

Selecting Particular Measures

To measure each outcome dimension, there are often a number of metrics or scales (e.g., the Medical Outcomes Study 36-Item Short-Form Health Survey [SF-36] or the Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC]) that can be utilized. Some metrics, such as the EuroQol Group 5-Dimension Self-Report Questionnaire (EQ-5D) scale to measure health-related quality of life, are generic metrics that can be used for multiple medical conditions. Other measures or scales are tailored to disease classes (e.g., SF-36 for orthopedics) or to individual medical conditions.

The particular measures chosen for each outcome dimension should reflect a number of considerations. First, measures should be selected that best capture the particular outcome from the perspective of the patient and medical science. Getting the measure right can have consequences. In in vitro fertilization (IVF), initial measurement focused on birth rates per IVF cycle, but this practice led to the implantation of numerous embryos and to a high number of multiple births (with a higher probability of complications). Over time, focus has shifted to birth rates per embryo implanted, and multiple births (especially triplet rates) have become a prominent outcome as well. The focus on measurement has played a major role in reducing triplet rates from 7 to 8% historically to less than 2%.

A second consideration in choosing measures is that, other things being equal, the selection of standard and tested measures will improve validity and enable comparison across providers. Third, measures should minimize ambiguity and judgment in scoring or interpreting, to ensure accuracy and consistency. Fourth, patient surveys should be utilized to measure outcomes such as functional status and discomfort that reflect patients' realities and are difficult for outside parties to measure. Here, standardized scales such as the SF-36 or the Beck Depression Index are preferable when available. Compromises will often be necessary in measure selection, but the measures chosen can be improved over time.

Many outcome measures can be tracked at various times in the cycle of care or cover periods of varying durations. For example, as noted above, the time to recovery can be disaggregated into the time to diagnosis and treatment plan, the time between diagnosis and treatment, and the elapsed time during treatment itself. Timing and duration should reflect relevance to patients as well as periods long enough to reveal results.

Practical considerations, such as the availability of data and cost of information gathering, will also play a role in the measures selected. For example, billing data are often more easily accessible than data from chart reviews or new data entry, and measures calculated from billing data can be the place to start as information systems are improved. Practical considerations may also influence the number and duration of measurement periods chosen. For most conditions, immediate complications are far easier to track than longer-term measures that require patient follow-up. Overall, however, the orientation should be on reducing the cost of capturing the right measures rather than limiting measures to those that are easy to obtain.

Developments in electronic medical records are already making outcomes far less costly to measure. Information technology infrastructure should be designed to facilitate the extraction of clinical data for measurement purposes, in addition to supporting the care delivery process.

Relationships among Outcome Dimensions

The relative importance of particular outcome dimensions can vary according to individual patient preferences, as noted above. For example, the ability to restore full physical activity may be especially important to an avid athlete or to someone whose employment involves physical labor.

Measurement of the hierarchy can reveal that levels are mutually dependent, as represented in the figures by the bidirectional arrows between levels. Progress at one level sometimes positively affects other levels, reflecting complementarities among outcome dimensions. For example, reducing complications or eliminating errors will not only reduce the disutility of care but speed up recovery.

Such complementarities among outcome dimensions reveal important leverage points for care improvement. For example, error reduction can have special significance beyond its direct Tier 2 benefits because errors may have cascading consequences for recovery, time, discomfort, and risk of recurrence. Error reduction, then, has been a strategic type of outcome improvement to focus on.

Cycle time is another particularly leveraged outcome dimension for value improvement. As discussed, cycle time is an outcome itself, reflecting the duration of anxiety, discomfort, and poor health for the patient. However, speeding up diagnosis and treatment (e.g., avoiding interruptions in care) and better managing complications and rehabilitation often have major benefits for the likelihood and degree of recovery as well as its sustainability, such as in cancer care. The value benefits (outcomes achieved per cost incurred) of cycle time are amplified by its impact on cost. Faster cycle time usually means that fewer resources are required to care for the patient. Cycle time, then, is an outcome dimension that every provider should measure and work to improve, though few have yet begun to do so. Avoidable complications are another important set of outcome dimensions with important complementarity and cost effects.

Measurement of the hierarchy can also make explicit the tradeoffs among outcome dimensions. For example, achieving more complete recovery may require more arduous or time-consuming treatment or confer a higher risk of complications. Mapping these outcome tradeoffs, and seeking ways to reduce them, is an essential part of the care innovation process.

In cases where there are tradeoffs among outcome dimensions, patients may place different weights on each level and dimension of the outcome hierarchy. The discomfort of treatment willingly endured may be affected, for example, by the degree of recovery possible. The long-term sustainability of recovery, such as 20-year implant survival for patients who undergo hip replacement, may matter less to older patients than the degree and speed of recovery. Or considerations of disfigurement may weigh heavily against the risk of recurrence — for example, when determining the amount of the breast to be resected from a patient with breast cancer.

Differences in the value patients place on individual outcome dimensions does not reduce the need to measure the full hierarchy but makes it more important to do so. Patients, their families, and their physicians, armed with information on a full set of outcomes, will be in a position to gain access to the treatments and providers that are best equipped to meet their particular needs.^{9,10} This level of outcome information goes well beyond what is currently available or even contemplated by medical societies and health plans in terms of consumer engagement.

Adjusting for Risk

The outcomes that are achievable will depend to some degree on each patient's initial conditions, sometimes also termed risk factors. Measuring and adjusting for initial conditions is therefore a crucial step in interpreting, comparing, and improving outcomes. In the case of breast cancer, for example, relevant initial conditions include the stage of disease at the initiation of care, the type of cancer (e.g., tubular, medullary, lobular, etc.), estrogen and progesterone receptor status (positive or negative), sites of metastases, and psychological factors, among others. Patients' compliance with treatment can also be interpreted as a risk factor — another reason why measurement of patient compliance is essential.¹

Risk adjustment is a complex topic, but I offer a number of strategic principles here. An illustrative set of initial conditions for breast cancer is shown in Figure 4. Initial conditions can affect all levels of the outcome hierarchy. Different initial conditions will often affect different outcome dimensions.

In order to evaluate outcomes for a medical condition, and especially to compare sets of outcomes over time or across providers, outcomes must be risk-adjusted or stratified by patient population based on the salient initial conditions. If initial conditions are not adjusted for, misleading conclusions can be drawn about the effectiveness of a treatment or provider that could mitigate the very purpose of outcome measurement. (An example of the risks of using outcome data without appropriate risk adjustment occurred when the state of Maine began to require drug-rehabilitation clinics to publish their outcomes. Subsequent studies have shown that the improvement in outcomes achieved in the years following the legislation were almost entirely attributable to clinics' turning away patients deemed likely to be problematic in order to increase their success rates.¹¹) Several efforts to gather and report outcomes have failed due to inadequate risk adjustment, which has led to resistance and rejection by the medical community.⁸ That said, there are a growing number of successful risk-adjustment approaches that confirm its feasibility and impact.

Adjusting for risk is not only necessary for measuring outcomes accurately, but also for improving them. Understanding the link between risk factors and specific patient health outcomes is critical for care decisions.

Finally, risk adjustment is not only important for making comparisons, but is also essential to mitigating the risk that providers or health plans will “cherry pick” healthier patients to improve measured outcomes. Inadequate risk-adjustment methods, as well as poor understanding of actual costs, are root causes of the underpayment of providers for handling patients with more complex conditions, both in the United States and elsewhere.¹ Flawed reimbursement for complex cases has many adverse consequences for value, ranging from inadequate care to excessive fragmentation of services as every provider is motivated to seek out “profitable” service lines and patient groups. Rigorous risk adjustment, coupled with corresponding reimbursement reform, will enable a move away from the current system of “profitable” and “unprofitable” interventions and patient populations and toward a system that encourages providers and health plans to focus on their areas of excellence.

Adjusting for initial conditions or risk normally involves two principal approaches. One is to stratify patient groups on the basis of the most important risk factors to allow outcomes for similar patients to be compared. This method is used in the area of in vitro fertilization, for example, where the Center for Disease Control reports birth rates according to maternal age cohorts and use of fresh or frozen embryos.

The other approach to risk adjustment is to utilize regression analysis to calculate expected outcomes, controlling for important patient risk factors. This allows average outcomes from different providers and periods to be adjusted for the patient mix or to be compared to expected outcomes for their particular patient populations. This method is utilized for outcome reporting in U.S. organ transplantation and in the Helios/AOK methodology in Germany focused on expected mortality for a wide array of medical conditions.⁴

Both stratification and risk adjustment depend on having sufficiently large patient populations to support statistically meaningful comparisons. To accumulate adequate numbers of patients, it may be necessary to aggregate patients over time or to examine outcomes for teams rather than for individual practitioners. In U.S. organ transplantation, for example, data are normally reported for 3-year periods. In in vitro fertilization, one of the weaknesses in the current reporting system is that results are reported only for patients in the most recent year, not over longer periods.

However, statistical power should not be the principal objective or driver of outcome measurement. The principal benefit of outcome measurement is to inform and stimulate practice improvement. The measurement and tracking of outcomes have major benefits even if the number of patients does not allow fine comparisons. In organ transplantation, for example, only a subset of centers has outcomes that are statistically better or worse than expected. However, all centers track their progress, and centers with weaker outcomes work actively to improve them. I will discuss the difference between outcome measurement and traditional clinical trials further below.

The challenge of risk measurement has often been used as an argument against outcome measurement. Although adjusting for risk is surely challenging in some cases and will never be perfect, there is ample evidence that doing so is feasible and that inappropriate comparisons among providers can be minimized.¹² Proven and accepted

risk-adjustment methods for complex fields already exist in the United States and several other countries. There is also no doubt that risk-stratification and adjustment methods will continue to improve with experience and that gaming of measurement will be mitigated over time.⁸

Risk Adjustment and Delivery Improvement

Even in its current imperfect state, risk adjustment is an essential tool for improving care delivery. Understanding and measuring patients' relevant initial conditions and their relationship to outcomes is indispensable to revealing new knowledge about medical conditions and their care.

The influence of initial conditions is partly inevitable — for example, the age of the mother appears to be a fundamental biologic influence on outcomes for in vitro fertilization.¹³ However, the influence of patient circumstances is partly a reflection of the state of understanding of a medical condition and its treatment. As clinical knowledge improves, certain risk factors may no longer meaningfully affect the outcomes of care, even though they may continue to influence the care process.

In vitro fertilization illustrates this learning process. Here, the biologic influences of age have been shown to weigh more heavily on egg production than on the ability to have a successful pregnancy. Through the use of donor eggs and improved technology for freezing a woman's own eggs, for example, older mothers are increasingly able to give birth to healthy children. So the impact of a mother's age has changed in terms of risk adjustment for the medical condition of infertility.

As learning occurs, risk adjustment for some initial conditions will become less necessary or even unnecessary for outcome comparison as providers manage them better. At the same time, new risk factors can emerge as sophistication in understanding a disease and in care delivery increases. This process of understanding and dealing with risk factors, then, is fundamental to driving value improvement. Advances in knowledge will reveal new, and perhaps more fundamental, initial conditions, such as genetic makeup. Yet improvements in care delivery over time can transform even genetic makeup from a risk factor to be adjusted for in comparing outcomes to a patient attribute that determines the best approach to successful care. Without systematic measurement of outcomes and risk factors, however, outcome improvement is hit-or-miss. The process of outcome measurement and risk adjustment is not only or even principally about comparing providers, then, but about enabling innovation in care.

These considerations suggest that it is preferable to err on the side of measuring more initial conditions rather than less and to create an explicit process for gradually revising the set of initial conditions used for risk adjustment. Most of all, the number and breadth of risk-adjustment studies and associated data collection must expand in every area of medicine to accelerate the rate of learning about care delivery.

The Outcomes Hierarchy and the Process of Value Improvement

Value improvement starts with defining and measuring the total set of outcomes for a medical condition and determining the major risk factors. Innovation in care delivery comes not only from focusing on individual outcome dimensions, but harnessing complementarities among various aspects of quality and reducing tradeoffs among outcome dimensions.

In medicine, as in most fields, progress in improving outcomes and value will be iterative and evolving. The outcomes hierarchy emphasizes that the pace of progress can vary across levels, and also among outcomes at a given level. As survival rates get high, for example, attention can shift to the speed and discomfort of treatment. Once the degree of recovery reaches an acceptable level, focus can shift to reducing tradeoffs between recovery and the risk of complications or care-induced illness, as in cancer therapy. Measurement of the entire outcome hierarchy not only encourages such improvements, but makes them more systematic and transparent.

Measuring the full hierarchy not only highlights multiple quality dimensions for improvement, but also expands the areas in which providers can distinguish themselves. As noted earlier, providers may achieve parity on some dimensions and then have to look to other dimensions to distinguish themselves. Or providers can concentrate on certain outcome dimensions that are weighted heavily by particular groups of patients.

In order to drive innovations in care, outcomes should be measured continuously for every patient, not just retrospectively in the context of discrete studies or evaluations. Whenever possible, outcomes should be measured in the line of care and inform continuous learning. The current approach to outcome measurement is skewed toward retrospective clinical studies, usually focused on a single end point. This bias towards clinical study methods is one of the reasons that outcome measurement remains so limited, despite its overwhelming benefits.

Comprehensive outcome measurement will enable a new type of clinical research, which focuses on overall care instead of controlled experiments around single interventions. Patient care is inevitably multidimensional, and actual care requires simultaneous choices on multiple variables and among numerous options. Conventional statistical methods need to be supplemented by careful study by clinical teams of patient-specific successes and failures. This kind of analysis seeks to identify common problems that arise, to discern patterns, and to develop hypotheses that give rise to learning, innovation, and further study.

Outcome Improvement and Cost Reduction

A major challenge in any field is to improve efficiency, and this is especially urgent in health care. One of the most powerful tools for reducing costs is improving quality, and outcome measurement is fundamental to improving the efficiency of care. Measuring the full outcome hierarchy provides a powerful tool for cost improvement that has been all but absent in the field. Comprehensive measurement of outcomes provides the evidence that will finally permit evaluation of whether care is actually benefitting patients and which treatments are most effective for each medical condition.

Historically, the overwhelming attention in outcome measurement has been directed at Tier 1 (health status achieved), particularly survival or mortality rates. At Tier 1, achieving better outcomes may (though by no means always does) require higher expenditures, especially when a new and expensive treatment or technology represents the only effective therapy. Such cases have led many observers to claim that innovation and new technology drive up health care costs. However, broader measurement of Tier 1 outcomes, notably functional status, will often open up opportunities for cost reduction. Improving the ability to function independently or return to work has huge cost consequences for the system.

Moreover, improvements in Tier 2 (process of recovery) and Tier 3 (health sustainability) outcomes almost invariably lower cost. Faster cycle time, fewer complications, and fewer failed therapies, for example, will have huge costs consequences. Tier 2 and 3 improvements can also reduce the cost of improving Tier 1 outcomes, because of the complementarities previously noted. For example, speeding up cycle time can also lead to more complete recovery, as is the case in cancer. Opportunities for dramatic improvement in Tier 2 and 3 outcomes engender great optimism for future cost containment; these opportunities have been overlooked because outcomes at these levels have been largely unmeasured and ignored.

Over the past several decades, joint replacement, new cancer therapies, organ transplantation, and many other new therapies were developed. In parallel, advancements in testing and diagnostic methods have allowed previously hidden conditions to be discovered or revealed much earlier. This stage of innovation, involving the development of new therapies for previously untreatable conditions and the discovery of previously hidden conditions, will almost inevitably raise cost, at least initially.

Today, however, the opportunity is different. Advancements in medical science have led to therapies that address most medical conditions in some way, albeit imperfectly. There will continue to be new tests and therapies where there were none before. However, the more common opportunity will be to drive dramatic value improvement in existing diagnostics and therapies, as well as to develop new, higher-value therapies that address diseases at earlier stages or more fundamental levels. A new era of rapid improvement in value in health care is possible. Comprehensive outcome and cost measurement, together with supporting changes in care organization, reimbursement, and market competition, will be needed to unlock and drive such value-based innovation.

Improving Value versus Rationing Care

Measuring the outcome hierarchy for each medical condition (and patient population receiving primary and preventive care) is indispensable for informing outcome improvement, assessing the value of alternative treatment approaches, and finding ways to deliver better outcomes more efficiently. Comparative-effectiveness research, in its present form, is important but not sufficient. It focuses largely on single interventions in highly controlled settings and sometimes incorporates just a single outcome or narrow set of outcomes. The outcome hierarchy is an important foundation for broadening and enriching clinical and comparative-effectiveness research at the medical condition level, as I have discussed. There have been efforts to monetize outcomes for purposes of calculating a benefit–cost ratio for alternative treatments. However, many such efforts tend to focus only on survival, even though survival is always one of a broader set of outcomes that matter to patients. Even for survival, assigning a monetary value is fraught with complexity, not to mention ethical issues. Is job productivity or earning power really a sufficient way to compare the health benefits of care, for example? Monetizing other important outcomes in the hierarchy from a benefit standpoint is even more challenging. For example, how should we value restoring the appearance of a patient with cancer or preserving a patient’s normal voice?

The use of quality-adjusted life-years (QALYs) or disability-adjusted life-years (DALYs) represents a broader approach to collapsing outcomes into a single measure.

Such measures embody a weighting of life expectancy based on quality of life. Quality of life is collapsed into a single number, determined using a variety of methods, despite the fact that it is inherently multidimensional and the relevant dimensions vary by medical condition.

At the medical condition level, we believe that there is little justification for shortcuts in measuring outcomes in driving value improvement. The full hierarchy of important outcomes needs to be measured and compared to cost. In evaluating alternative care delivery approaches, the task is to examine how the set of outcomes improves, and how improvement in the set of outcomes relates to cost. If one or more outcomes in the hierarchy improve while others remain stable, the set of outcomes improves. Value improves if outcomes improve at equal or lower cost, or if outcomes are stable at meaningfully lower cost.

There is no benefit to collapsing or suppressing outcome dimensions in making this evaluation at the medical condition level — quite the contrary. All parts of the outcome hierarchy are important to patients, and progress on each dimension is beneficial. Examinations of Tier 2 and Tier 3 outcomes, which are rarely considered in comparative-effectiveness studies, are powerful tools not only for outcome improvement but also cost reduction. There are certainly cases of tradeoffs — in which better outcomes occur only at much higher costs. However, there are virtually unlimited opportunities for improvement in the outcome hierarchy that do not involve such tradeoffs, and this is where attention in care improvement should be focused.

Monetization of outcomes and QALYs or DALYs are often used to compare the value of care across medical conditions. We know that for each medical condition, the set of relevant outcomes will be different. QALYs and DALYs focus just on those outcomes that can be readily standardized — again, survival and certain generic aspects of quality of life. Once again, the validity and comparability across conditions of these measures is highly questionable.

This effort to standardize and collapse outcomes to a single measure also suffers from a deeper problem. The whole approach assumes that the value of care for each medical condition is fixed and that care must be rationed. Optimizing within fixed constraints comes naturally to some economists but has proven shortsighted time and time again. In a field where outcomes are all but unmeasured, and where cost is poorly understood, there are major opportunities to improve outcome and value in the care for every medical condition. This is where the field should focus. Setting policies to enable and incentivize innovation should be our approach, rather than assuming that the value is fixed and focusing on choosing which patients should receive care. Given the major improvements in outcomes and efficiency observed in areas where there has been rigorous outcome measurement, there is every reason to hope that rationing will not be necessary except in extreme cases.

Health care is on a dangerous path if the primary rationale for outcome measurement is rationing of care rather than outcome and value improvement. Standardized outcome-measurement approaches will not well serve the needs of improving clinical practice, and they will disenfranchise providers. Turning to rationing without taking aggressive steps toward improving outcome and efficiency is a failure of policy — and will also prove unacceptable to patients and their families. Moreover, such policy will fail to be implemented when political realities intrude.

Conclusion

Outcome measurement is the single most important tool to drive innovation in health care delivery. The feasibility, practicality, and impact of outcome measurement have been conclusively demonstrated. Every provider can begin to measure the outcomes hierarchy in the medical conditions it serves, and track its progress versus past performance. Outcome measurement can begin for a subset of medical conditions and expand over time as infrastructure and experience grow.

This article provides a framework for systematically identifying the full set of outcomes for each medical condition, exploring the relationships among them, and revealing risk factors. Today, numerous voluntary and mandatory programs track different measures for subsets of providers, payers, and patient populations. The challenge is to make outcome measurement ubiquitous and an integral part of health care delivery.

Over time, the goal should be to establish uniform national and international outcome-measurement standards and methods. The feasibility of such standards has been conclusively demonstrated. Rather than resting with today's consensus organizations or government entities that are caught up in politics, responsibility for outcome measurement standards should be delegated to a respected independent organization, such as a new affiliate of the Institute of Medicine. Measurement and reporting of outcomes should eventually become mandatory for every provider and health plan. Reporting by health plans of health outcomes for its members, according to medical condition and patient population, using data drawn from providers' reporting, will help to shift health plans' focus from short-term cost reduction to value improvement.

As comprehensive outcome measurement is being phased in, every provider should report experience (i.e., the volume of patients treated for each medical condition), along with the procedures and treatment approaches utilized. Experience reporting will begin to help patients, their doctors, and health plans find the providers with the expertise that meets their needs. It will also highlight the fragmentation of care across facilities and providers and inform a rationalization of service lines. The most important users of outcome measurement are providers, for whom comprehensive measurement will lead to substantial improvement.⁵ The most important purpose of outcome measurement is improvement in care, not keeping score. Outcome measurement is also a powerful vehicle for bringing teams together and improving collaboration in a fragmented field. There is much evidence that the very act of measuring outcomes leads to substantial improvement. Public reporting of outcomes is not necessary in order to reap important benefits, and studies have revealed that confidential, internal reviews can motivate providers to improve their performance.¹⁴ Public reporting must be phased in carefully to win provider confidence. However, eventual progression to public reporting will accelerate innovation by further motivating providers to improve and permitting all stakeholders to benefit fully from outcome information.

From Harvard Business School, Boston.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

This article draws on Redefining Health Care and subsequent research. I thank Jennifer F. Baron, Carolyn Daly, and Rasmus Molander for this research; and Andrew Huang, Sachin Jain, Christopher Lis, Joan Magretta, Michael McGinnis, Margaret O’Kane, Elizabeth Teisberg, Scott Wallace, and Jason Wang for helpful comments.

References

1. Porter ME. What is value in health care? *N Engl J Med* 2010; 363:2477-81 (DOI: 10.1056/NEJMmp1011024).
2. Porter ME. Value in health care. *N Engl J Med* 2010.
3. Porter ME. Defining and introducing value in health care. Evidence-Based Medicine and the Changing Nature of Health Care: 2007 IOM Annual Meeting Summary (2008):161-72.
4. Porter ME, Molander R. Outcomes measurement: learning from international experiences. Institute for Strategy and Competitiveness, Harvard University Working Paper. <http://www.hbs.edu/rhc/prior.html>. 2010.
5. Porter ME, Baron JF, Chacko JM, Tang R. The UCLA Medical Center: Kidney transplantation," HBS No. 711-410 (Boston: Harvard Business School Publishing, 2010).
6. Porter ME, Rahim S, Tsai B. In-vitro fertilization: outcomes measurement. HBS No. 709-403 (Boston: Harvard Business School Publishing, 2008).
7. Feely TW, Fly HS, Albright H, Walters R, Burke TW. A Method for Defining Value in Healthcare Using Cancer Care as a Model. *Journal of Healthcare Management*. 55:6 November/December 2010: 399-412.
8. Porter ME, Teisberg EO. Redefining Health Care. Boston, MA: Harvard Business School Press, 2006.
9. Wennberg JE. Dealing with medical practice variations: a proposal for action. *Health Affairs (Millwood)* 1984;3(2):6-32.
10. Wennberg JE. Improving the medical decision-making process. *Health Aff (Millwood)*.1988;7(1):99-106.
11. Shen Y. Selection incentives in a performance-based contracting system. *Health Services Research*, April 1, 2003.
12. Boyce N. Using outcome data to measure quality in health care. *International Journal for Quality in Health Care*. 1996;8(2):101-4.
13. Centers for Disease Control and Prevention. 2004 Assisted reproductive technology success rates: national summary and fertility clinic report. <http://www.cdc.gov/ART/ART2004/index.htm>. Updated April 2007. Accessed October 29, 2008.
14. Antonacci AA, Lam S, Lavarias V, Homel P, Eavey RA. A report card system using error profile analysis and concurrent morbidity and mortality review: surgical outcome analysis, part II," *Journal of Surgical Research*, Vol. 153, No. 1, May 1, 2009

Figure 1. The Outcome Measures Hierarchy.

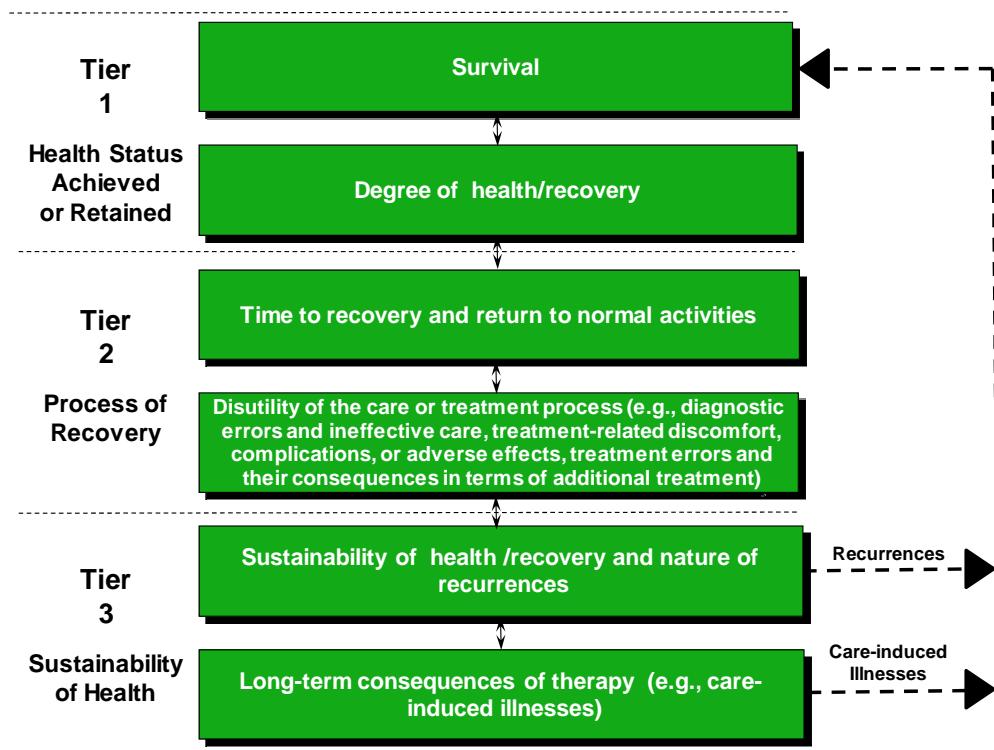


Figure 2. Illustrative Outcome Hierarchy for Breast Cancer and Knee Osteoarthritis.

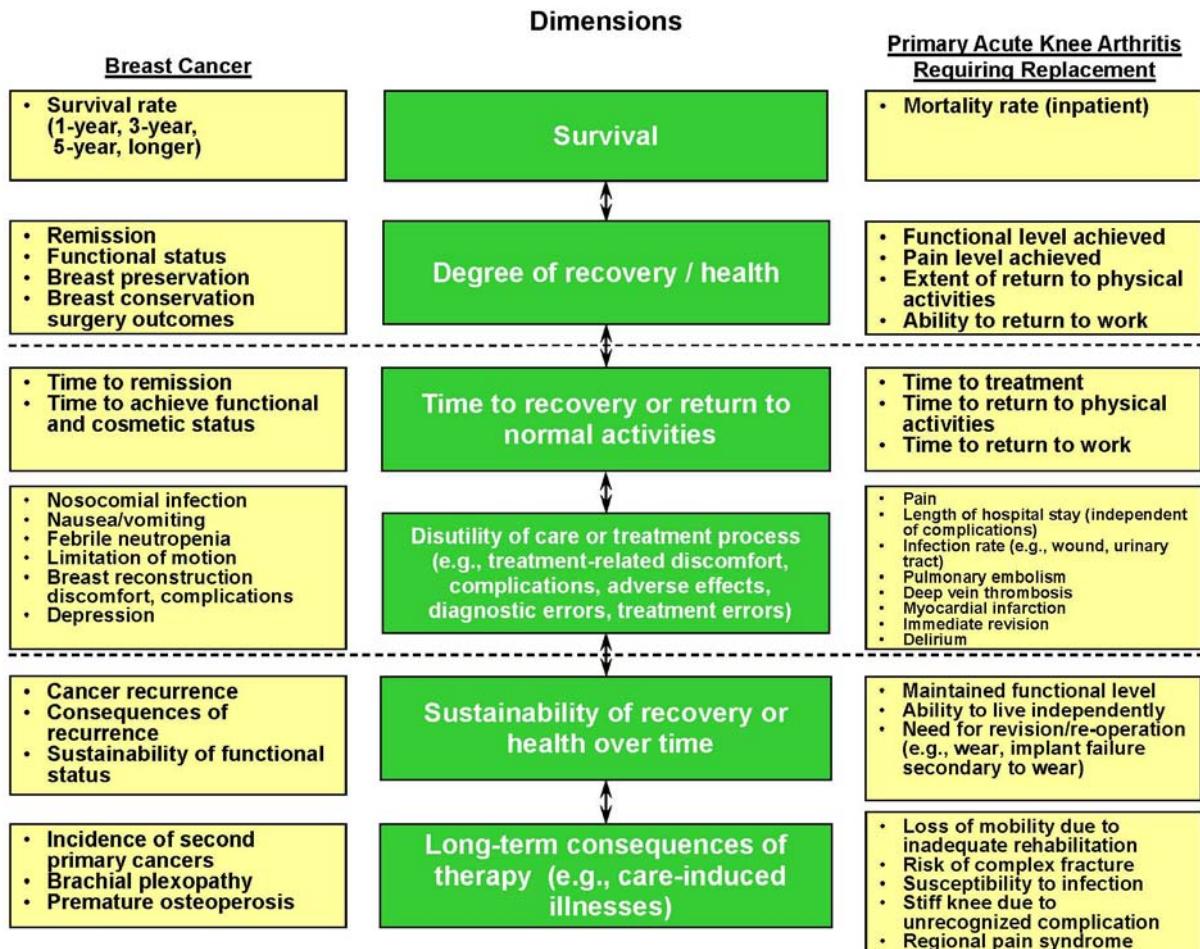


Figure 3. The Care Delivery Value Chain (Breast Cancer).

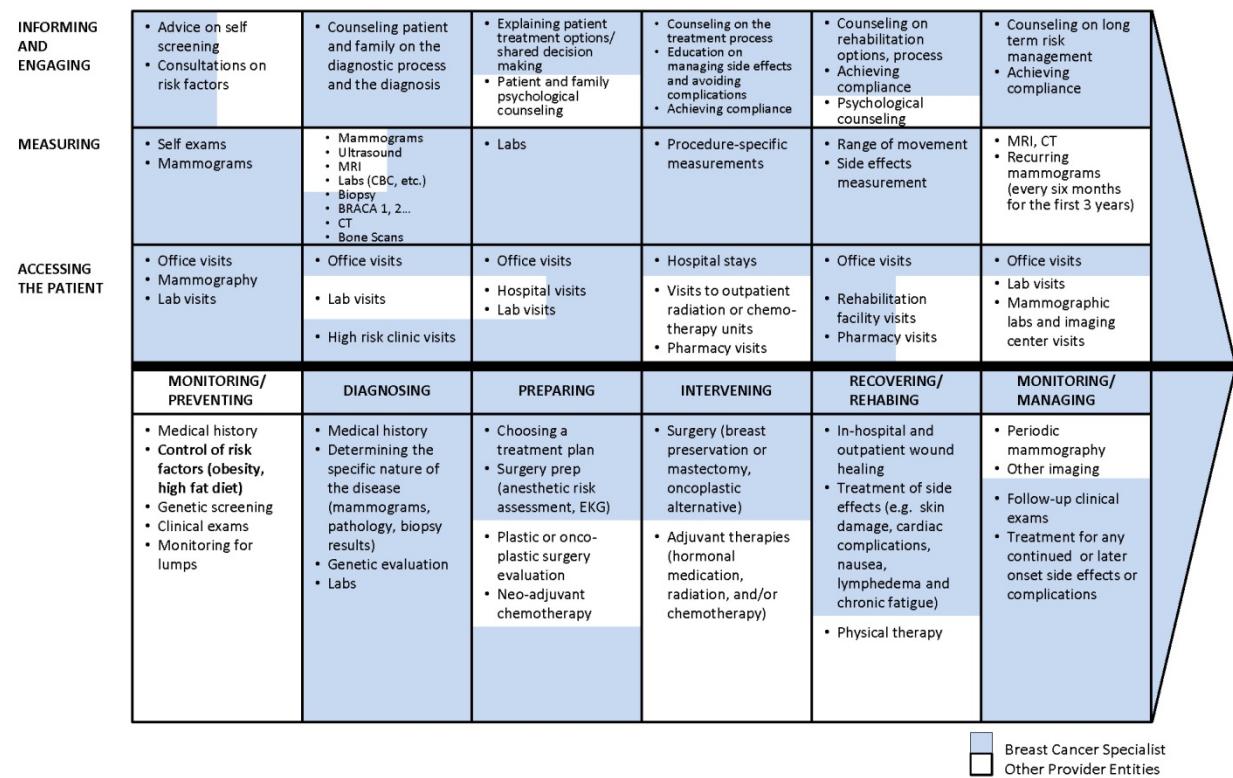


Figure 4. Illustrative Risk Factors for a Patient with Breast Cancer.

- Stage of disease
- Type of cancer (infiltrating ductal carcinoma, tubular, medullary, lobular, etc.)
- Estrogen and progesterone receptor status (positive or negative)
- Sites of metastases
- Previous treatments
- Age
- Menopausal status
- General health, including co-morbidities
- Psychological and social factors