



Quality Reporting on Medicare's *Compare* Sites

Lessons Learned from
Consumer Research, 2001–2014

DECEMBER 2015



MATHEMATICA
Policy Research



Prepared for:

David Miranda, PhD
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore MD 21244

Prepared by:

L&M Policy Research
1743 Connecticut Avenue, NW, Suite 200
Washington DC 20009

Mathematica Policy Research
600 Alexander Park, Suite 100
Princeton NJ 08540



L & M
POLICY
RESEARCH

MATHEMATICA
Policy Research

Acknowledgements

Funding for this report was provided by the Centers for Medicare & Medicaid Services (CMS), under contract 500-2011-000191, HHSM-500-T003. The report is a culmination of several years of research conducted by team members from L&M Policy Research (L&M) and Mathematica Policy Research. Contributing authors include Margaret Gerteis, Ph.D. and Cicely Thomas, M.Sc., from Mathematica and Lauren Blatt, Sally Crelia, M.P.H., Alyson Marano Ward, M.P.H., R.N., Kelly Moriarty, Monica Sarmiento, M.S., P.M.P, Myra Tanamor, M.P.P, and Russ Tisinger, Ph.D., from L&M.

The authors would like to recognize Eugenia Mattison-Gibson, M.B.A., and Memuna Ifedirah, from CMS and Rosemary Lee, R.N. (formerly with CMS), for their ongoing assistance. The team would also like to thank the various members of the *Hospital Compare* and *Home Health Compare* stakeholder workgroups who generously gave their perspective and insight.

Finally, we would like to especially thank David Miranda, Ph.D., for his continuous support and guidance throughout the course of this collaborative endeavour.

Table of Contents

I.	Introduction	1
	Public Reporting on Medicare.gov	1
	A Social Marketing Approach to Consumer Research	2
	Lessons Learned from Consumer Research.....	4
II.	Understanding the Context of Consumers' Choices and Decisions	5
	Selecting a Health Plan	5
	Selecting a Physician	6
	Selecting a Hospital	7
	Arranging Home Health Services.....	8
	Nursing Home Placements	9
	Choosing a Dialysis Facility	10
III.	Professionals' vs. Consumers' Understanding of Quality and Efficiency	12
	Measuring and Reporting on Quality and Efficiency: The Professional Perspective	12
	Consumers' Take on Publicly Reported Measures.....	13
	How Consumers Think About Quality	13
	Defining a Common Framework for Communicating About Quality.....	15
IV.	Conveying Performance Information	17
	Establishing the Context: What is This and Why is it Important?	18
	Organizing and Layering Information.....	18
	Explaining What is Being Measured and How it Relates to Quality.....	19
	Understanding the Metrics	22
	Displays: What Do Numbers Mean?.....	25
	Translation and Cultural Relevance.....	28
V.	Usability and Navigation	29
	Context	30
	Presentation and Display	31
	Ease of Use/Decision Support.....	31



I. Introduction

PUBLIC REPORTING ON MEDICARE.GOV

The Centers for Medicare & Medicaid Services (CMS) has been the driving force behind public reporting on health plan and provider performance in the United States for more than a decade. Based on the premise that making performance information available to the public will lead to a more rational health care marketplace, Medicare's *Compare* web tools were designed initially as sites for consumers to find and compare the performance of health plans, physicians, hospitals, home health agencies, nursing homes, and dialysis facilities. The Patient Protection and Affordable Care Act (ACA) of 2010 now directs CMS to further refine and enhance efforts to educate consumers about quality and empower them to manage their health care.

Over the years, CMS has begun to change its stance on public reporting, from a policy of posting data without interpretation to recognizing that consumers need at least some interpretative guidance (even if only to say whether higher or lower numbers are better). As the volume of publicly reported information grows, CMS is also beginning to support the idea of summarizing and synthesizing information through composites and symbols, acknowledging that consumers' ability to process page after page of statistics and bar graphs is limited.

However, the lay public is not the only intended audience. Public reporting is, increasingly, a vehicle for making performance information available to different constituencies for different purposes, and the *Compare* sites are also now the designated channels for publicly reporting performance information related to federal initiatives, including incentive payment programs for value-based purchasing, provider participation in quality-reporting registries, use of health information technology, and measures of resource use. This expanding mandate complicates the task of conveying information to consumers. Statutory time lines can leave little time to consider what the implications or

underlying messages are for lay users of the *Compare* tools, or to design new website content and displays for a consumer audience.

Since CMS first began planning its quality initiatives and public reporting agenda nearly 15 years ago, the authors of this report have researched quality reporting issues from a consumer perspective under a number of contracts with CMS — from early formative research on consumers' information-seeking behavior, decision-making scenarios, and target audiences, to the development and testing of prototype *Compare* tools, cognitive testing of new measures and displays, and research on alternative reporting formats and organizational frameworks.

This report draws on that body of research. What differentiates this work from that of consumer researchers and consultants operating in other venues is our specific focus on CMS's quality initiatives and *Compare* tools and the constraints of public reporting in that context. We recognize that CMS's decisions are informed not just by consumer research, but also by policy judgments and politics, by technological and financial limitations, and by competing stakeholder interests — issues that will always be in play. Our objective, then, is not to create a primer for public reporting (which already exist¹), or to make specific recommendations that may or may not be feasible to implement. Rather, our aim is to offer some insights, based on what we have learned, about the interplay of factors that shape consumers' perceptions and ability to engage with publicly reported information. These issues, too, must be taken into account in the ongoing development of CMS's *Compare* tools.

1 See, for example, <https://cabps.ahrq.gov/consumer-reporting/talkingquality/>, sponsored by the Agency for Healthcare Research and Quality.

A SOCIAL MARKETING APPROACH TO CONSUMER RESEARCH

Since we first began working with CMS to help design and promote its quality-reporting initiatives and *Compare* websites, social marketing principles have provided the framework for our consumer research. The “four P’s” of marketing emphasize the design of a *product* that meets an identified need for an identified consumer segment, made available in a *place* that is accessible to consumers when they need it, at a *price* they are willing to pay, and product *promotion* that uses the right messaging to convey this information to the right audience. Social marketing applies these principles to the development and promotion of products designed to serve a social purpose, rather than for commercial gain. In this case, the social purpose, broadly speaking, is to improve the quality of health care and to engage consumers in using publicly reported information to that end.

Within the constraints of CMS’s developmental time lines and budgets for the quality initiatives and web tools, the research supporting this work has followed a defined arc, beginning with formative research to understand who the consumers are and the context in which they make their decisions — what factors enter into their choices, who and what influence them, what constraints they face. This is the first step in defining the primary audience for publicly reported information, as well as secondary audiences who influence their actions or decisions. To this end, much of the early work we undertook with CMS explored consumers’ information-seeking and decision-making behavior under different scenarios — focusing first on planning for post-acute or long-term care (in nursing homes or home health settings) to inform CMS’s Nursing Home and Home Health Quality Initiatives and focusing later on acute care in hospital settings to support the development of *Hospital Compare*.² This work relied heavily on small group discussions with patients and caregivers, health care providers, and professionals who serve as information intermediaries for consumers.

Once the context of decision making is better understood, the research focus shifts to product design, product positioning, and messaging — that is, the design of the *Compare* tools, themselves. For this work, we have more often used one-on-one, in-depth interviews with primary or secondary consumer audiences, both to develop the content and format of the tools and to test their usability. Because the price or “cost” of the product, in this case, is related to the cognitive burden of using the information, we

² Formative and developmental research for CMS’s *Dialysis Facility Compare* was conducted by RTI International.

SOCIAL MARKETING

SOCIAL MARKETING applies the principles of product marketing, often described in terms of the “four P’s,” to the advancement of social agendas:

- **PRODUCT:** Does the product (the publicly reported health services information) meet consumers’ needs?
- **PLACE:** Is the health services information accessible to consumers when and where they need it?
- **PRICE:** Is the cost of the product (the time and effort it takes to access the information) worth the benefit to consumers?
- **PROMOTION:** How can consumers learn about the information on the CMS website?

have used cognitive testing techniques to determine how readily and accurately consumers can understand and interpret narrative content and data displays in different formats. We have used this approach both to inform the initial development of the *Compare* tools and to test ways to incorporate changes and additions, as new measures of quality and efficiency are added.

Although much of our focus has been on formative research, product development, and refinement, we have also explored ways to get publicly reported information into consumers’ hands at key points in the decision-making process, as well as the potential use of decision aids, checklists, and other tools to make the information more usable. For example, when research on post-acute care placement decisions shed light on the central role that hospital discharge planners play, we explored the feasibility of using these professionals as intermediaries to disseminate information on the quality performance of nursing homes or home health agencies to consumers. Our early research on nursing home placement decisions also revealed that although family caregivers wanted to visit a nursing home before placing a family member there, they did not know what to look for

or what questions to ask. We therefore developed and tested a checklist that would link the publicly reported quality data for a given nursing home to observable practices or questions to ask.

As suggested above, the research methodologies we have used have varied somewhat, depending on the purpose of the task at hand and the audience of interest. Research in a new topic area typically begins with a review of the research literature and an environmental scan. This may include reviewing existing research on health care communication, risk communication, decision making, health literacy, and other factors, as well as public reporting practices in other venues. Key informant interviews help the research team better understand the rationale and context for reporting specific measures or measurement methodologies. Reviewing research questions and methods with groups of stakeholders can provide additional insights and perspectives. Direct observation can also be a useful qualitative research technique — for example, shadowing hospital discharge planners during the workday to better understand the environment and context in which post-acute care placement decisions are made.

For primary research with consumers, family caregivers, health care providers, and other professionals, we have most often used two common qualitative research methodologies — focus groups and in-depth interviews — each of which has its uses and limitations:

- **FOCUS GROUPS** are facilitated group discussions appropriate for exploring values, attitudes, perceptions, and behaviors based on participants’ shared characteristics or common experiences, and are particularly useful in formative research. Because group dynamics can obscure important differences among individual group members, however, we have found focus groups to be less effective in product development or usability testing with consumers. Moreover, focus groups may not work well for physicians, whose need to establish their authority in a group of peers can preclude productive give-and-take discussion.
- **ONE-ON-ONE INTERVIEW METHODOLOGIES** are more appropriate for in-depth exploration of individual attitudes and experiences or potentially sensitive topics; cognitive testing of users’ understanding and interpretation of information; and usability testing to observe how individuals navigate a particular web tool, what features they use (or miss), and how they accomplish particular tasks. As noted above, one-on-one interviews are also the best method for eliciting information about the perceptions and experiences of physicians or other professionals who may feel the need to compete in focus group settings. However, one-on-one interviews do not

allow researchers to observe how participants would interact with others around a particular topic or piece of information.

- **DYADS** (groupings of two research participants with one interviewer) offer a useful alternative to one-on-one interviews, in some cases — for example, when exploring the role of family caregivers or other intermediaries on consumers’ understanding of information.

Regardless of the methodology, the aim always is to recruit research participants from the community who are broadly representative of the target audience of interest. In most cases, we have found that this is best accomplished by using community-based market research firms to handle recruitment, using screening protocols developed by the research team to achieve the desired participant mix. Relying on referrals from professionals or community-based organizations may yield a more skewed participant sample, we have found, although it may be necessary in some circumstances.

RESEARCH OBJECTIVES FOR MEDICARE’S COMPARE TOOLS

- **IDENTIFYING TARGET GROUPS AND TARGET AUDIENCES**, the people whose behavior you aim to influence (or those who influence them).
- **CONDUCTING FORMATIVE RESEARCH** to understand target audience needs, perceptions, and behaviors.
- **DEVELOPING COMMUNICATION STRATEGIES** to reach these audiences and influence their behavior.
- **DEVELOPING PRODUCTS**, and testing and refining them to tailor messaging and content to the strategies of the communication campaign.
- **TESTING USABILITY**, to determine whether, and how readily, the product can be used.

LESSONS LEARNED FROM CONSUMER RESEARCH

Most of the consumer research that the authors (and other researchers in the field) have conducted in recent years has focused on presenting comparative performance data to the lay public in ways that will inform their choice of health care plans or providers. However, there is scant evidence to suggest that consumers consult or use such information for this purpose³ — in part, because other issues loom larger in their personal decisions, in part because the performance information currently reported may not seem relevant to the decisions they face, and in part because they may not know such information exists. Informing consumer choice implies a focus on the individual and would require tailoring information to the specific decisions that individuals face, in practice.

With these considerations in mind, our research team has also begun to take a somewhat different tack in our approach to consumer research, framing discussions around questions about health care quality and perceived variations in quality, rather than around issues of choice, per se. This approach is based on the idea, supported by some evidence,⁴ that public reporting can stimulate improvements in quality and efficiency through mechanisms other than consumer choice. The objective (and challenge), in this case, is to engage consumers in thinking about health care quality and efficiency at a conceptual level that may or may not bear directly on the individual choices they face.

In this report, we synthesize key findings from both lines of inquiry. In Section II, we begin by summarizing what we have learned about consumer decision making under different circumstances. In Section III, we highlight similarities and differences between consumers' and professionals' understanding of quality and efficiency in health care. In Section IV, we discuss

the specific challenges of conveying clear, accurate, and usable performance information to consumers. In Section V, we present some of the common issues related to website navigation and usability that we have encountered.

Our aim is not to document in detail everything we have learned over the years, but to highlight the most prominent and consistent themes that seem to us most relevant to engaging consumers in questions about health care quality and efficiency. Nor do we seek to synthesize the work of other researchers in this field, although we regularly review this literature as part of our ongoing work. Unless otherwise noted, the observations offered here are based only on our own work.

ASSUMPTIONS UNDERLYING “INFORMED CHOICE”

- Consumers are the decision makers.
- Consumers have a choice.
- Consumers perceive consequences to their choices.
- Publicly reported information is understandable to consumers and relevant to their perceived needs.
- Consumers see publicly reported information as reflecting meaningful differences in performance.

³ This is supported by several recent systematic research reviews. See, in particular, Totten, A. M., J. Wagner, A. Tiwari, et al., *Closing the Quality Gap: Revisiting the State of the Science* (Vol. 5: Public Reporting as a Quality Improvement Strategy), Rockville MD: Agency for Healthcare Research and Quality, July 2012 (Evidence Reports/Technology Assessments, No. 208.5); and Fung, C. H., Y. W. Lim, S. Mattke, C. Damberg, and P. G. Shekelle, “Systematic Review: The Evidence That Publishing Patient Care Performance Data Improves Quality of Care,” *Annals of Internal Medicine*, 2008, 148(2):111–123.

⁴ See, for example, Lamb, G. C., M. A. Smith, W. B. Weeks, and C. Queram, “Publicly Reported Quality-Of-Care Measures Influenced Wisconsin Physician Groups to Improve Performance,” *Health Affairs*, 2013, 32(3):536–543; and Hibbard, J. H., J. Stockard, and M. Tusler, “Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts?” *Health Affairs*, 2003, 22(2):84–94.

II. Understanding the Context of Consumers' Choices and Decisions

One of the central premises of public reporting has been that comparative data on performance will lead to better quality and efficiency by informing consumer choice in a competitive marketplace. This assumes that consumers are the decision makers, have choices, perceive consequences to those choices, and can see meaningful differences in performance that are relevant to their decisions.

In practice, however, this may not be the case. In this chapter, we review common scenarios around decisions about selecting health plans and the types of providers included in each of Medicare's *Compare* tools, in light of assumptions underlying informed choice.

UNDERSTANDING THE DECISION PROCESS

- What is the context for the decision?
- What information do consumers look for or need?
- Who makes the decision?
- How much choice do they have?
- How does timing affect their ability to weigh choices?

SELECTING A HEALTH PLAN

<p>What is the context for the decision?</p>	<ul style="list-style-type: none"> • Consumers tend to look for a health plan only if they need to. If they already have a plan, they might not be motivated to review their options or take action unless there is some change in their life circumstances (for example, a move or a change in employment, health, financial, or marital status).
<p>What information do consumers look for or need?</p>	<ul style="list-style-type: none"> • Consumers look for information relevant to their particular needs and circumstances, including specific benefits and covered services, provider networks, out-of-pocket costs (for premiums and co-pays), and prescription drug coverage.
<p>Who makes the decision?</p>	<ul style="list-style-type: none"> • Consumers are usually the primary decision makers, but their decisions may be influenced by others (such as family members, friends, counselors). Formal and informal information intermediaries (such as State Health Insurance Assistance Program counselors) may also play a role.
<p>How much choice do they have?</p>	<ul style="list-style-type: none"> • All Medicare beneficiaries have some options, although choices may be limited by personal circumstances.
<p>How does timing affect their ability to weigh choices?</p>	<ul style="list-style-type: none"> • The timing of decisions, linked to eligibility or open-enrollment periods, is usually sufficient to allow consumers to gather, review, and compare information about health plan characteristics and performance to inform their decisions.

SELECTING A PHYSICIAN

<p>What is the context for the decision?</p>	<ul style="list-style-type: none"> • Circumstances surrounding a search for a physician vary, but may be affected by a change in patient’s life circumstances (for example, a move, change in health insurance, or a new health concern), a physician’s move or retirement, or a patient’s dissatisfaction with an existing physician.
<p>What information do consumers look for or need?</p>	<ul style="list-style-type: none"> • Consumers consider other physicians or health professionals, friends, and family members to be the most trusted sources of information about physicians. • They may also look for information about a physician’s credentials and specialized training or experience; hospital affiliation; personal characteristics (including gender, race or ethnicity, primary language spoken); proximity; participation in the patient’s insurance plan; and availability.
<p>Who makes the decision?</p>	<ul style="list-style-type: none"> • Consumers (or their family members) usually select their own <i>primary care physicians</i>. • Primary care or other physicians usually refer patients to <i>specialists</i>, although consumers may sometimes seek out specialists on their own (for example, for a second opinion).
<p>How much choice do they have?</p>	<ul style="list-style-type: none"> • Choices of both primary care and specialist physicians may be limited to the provider network covered by the consumer’s health plan, or by providers’ willingness to accept the patient’s insurance. • Choices will also be limited by the supply of physicians available in a given specialty in a particular geographic market.
<p>How does timing affect their ability to weigh choices?</p>	<ul style="list-style-type: none"> • Timing will vary with circumstances, depending on the urgency of the need.

SELECTING A HOSPITAL

<p>What is the context for the decision?</p>	<ul style="list-style-type: none"> • Except in emergency situations or for childbirth, patients typically do not select the hospital. They go to the hospital to which their doctor refers them. • Under emergency circumstances, consumers choose hospitals based on proximity, reputation, prior experience, or their doctor's recommendation.
<p>What information do consumers look for or need?</p>	<ul style="list-style-type: none"> • For emergencies, consumers are interested in information about wait times. • For non-emergency admissions, consumers are most interested in information about the hospital's (and physician's) track record in treating their condition. • Consumers may also be interested in physician and other staff credentials, staffing ratios, and a hospital's recognized areas of expertise ("centers of excellence").
<p>Who makes the decision?</p>	<ul style="list-style-type: none"> • In emergency situations, patients, their friends or family members, or emergency medical personnel make decisions about which hospital to use. • In non-emergency situations, the choice of hospital is usually determined by the patient's doctor.
<p>How much choice do they have?</p>	<ul style="list-style-type: none"> • Hospital choice may also be limited by insurance coverage or the availability of needed specialized services, as well as the number and types of hospitals in the geographic market.
<p>How does timing affect their ability to weigh choices?</p>	<ul style="list-style-type: none"> • There is no time to weigh information and make deliberative choices in most emergency situations. • For childbirth or elective hospital admissions, patients and their family members may have the time and motivation to consider options and engage in informed decision making.

ARRANGING HOME HEALTH SERVICES

<p>What is the context for the decision?</p>	<ul style="list-style-type: none"> • Most consumers understand very little about Medicare-covered home health services. They think in terms of the therapist or nurse who will come to see them, rather than the agency that will send them.
<p>What information do consumers look for or need?</p>	<ul style="list-style-type: none"> • Patients and families need to know what specific services Medicare will and will not cover. They also need to know what services will be provided and how often and for how long they will be provided.
<p>Who makes the decision?</p>	<ul style="list-style-type: none"> • Most patients have no practical sense about what to look for in a home health agency. They willingly defer to the discharge planner's (or doctor's) judgment, unless they have a preference based on prior experience. • Arrangements for post-acute home health care are almost always made by hospital discharge planners, case managers, or physicians.
<p>How much choice do they have?</p>	<ul style="list-style-type: none"> • Although Medicare may require that patients be offered more than one agency to choose from, available options will be limited to agencies in the patient's geographic area that have the resources to provide the specific services needed when those services are required. • The availability of home health services and competition in home health markets also varies widely in different geographic regions.
<p>How does timing affect their ability to weigh choices?</p>	<ul style="list-style-type: none"> • Patients who have some scheduled procedures (such as orthopedic surgery) may know in advance that post-acute home health care will be required, and discharge planning might begin early. • However, arrangements for home health care are often made within a few hours of a patient's discharge, leaving little time for consumers to weigh options. • Hospital discharge planners are most interested in an agency's ability to provide timely and reliable care (especially for difficult cases), because they want to expedite a patient's discharge and prevent short-term readmission.

CHOOSING A DIALYSIS FACILITY

<p>What is the context for the decision?</p>	<ul style="list-style-type: none"> • Patients with end-stage renal disease may have little understanding of what dialysis will entail, or what to look for, until they have experienced it. • After they have been on dialysis for a while, patients are better able to understand, process, and consider comparative information on facility performance, and plan accordingly.
<p>What information do consumers look for or need?</p>	<ul style="list-style-type: none"> • Factors important to patients in choosing a facility initially include physician (nephrologist) affiliation; proximity to home or work and/or access to transportation; space availability and convenience of hours of operation; comfort and amenities; and cleanliness. • Unlike consumers facing most other choice decisions, patients who have had a long-term relationship with a dialysis facility have a deeper understanding of clinical processes of care and the associated measures of performance.
<p>Who makes the decision?</p>	<ul style="list-style-type: none"> • Patients with end-stage renal disease must select an outpatient facility for their ongoing dialysis care. Physicians (nephrologists), hospital discharge planners, or case managers may assist with initial placement decisions.
<p>How much choice do they have?</p>	<ul style="list-style-type: none"> • Available options will be limited by the number of facilities that have space available, are accessible to the patient, and have the capacity to provide the specific services required (hemodialysis, peritoneal dialysis, home dialysis instruction, home dialysis). • Facilities might also refuse to accept patients who have failed to keep scheduled appointments in the past.
<p>How does timing affect their ability to weigh choices?</p>	<ul style="list-style-type: none"> • Patients with chronic kidney disease may know in advance that they will require dialysis and be able to plan ahead, in collaboration with physicians or case managers. • However, many patients might be unaware, or in denial, until they are in crisis and require dialysis urgently.

NURSING HOME PLACEMENTS

<p>What is the context for the decision?</p>	<ul style="list-style-type: none"> • Decisions about nursing home placements are usually precipitated by a medical crisis, when emotions are running high. • Nursing homes have a negative connotation, and consumers worry about problems with quality and safety. • Family members may have a strong cultural aversion to the idea of placing a sick and elderly relative in a nursing home.
<p>What information do consumers look for or need?</p>	<ul style="list-style-type: none"> • Consumers express interest in information about staffing ratios and safety inspections. • They may resist performance information based on negative outcomes, however, because it reinforces their fears.
<p>Who makes the decision?</p>	<ul style="list-style-type: none"> • Decisions about post-acute, short-term, or long-term nursing home placements are almost always made by patients' family members, in consultation with hospital discharge planners, case managers, or physicians.
<p>How much choice do they have?</p>	<ul style="list-style-type: none"> • Options will be limited by the availability of beds and services at the appropriate level of care, family finances or insurance coverage, and market characteristics.
<p>How does timing affect their ability to weigh choices?</p>	<ul style="list-style-type: none"> • Most family members resist thinking about the potential need for a nursing home placement or other long-term care until circumstances compel them to do so. • Nursing home placement decisions typically take place within a few days of a pending discharge from a hospital stay, which may leave little time to explore options.

In sum, the circumstances surrounding the kinds of health care decisions Medicare's *Compare* tools are intended to inform vary markedly, and the assumptions underlying the concept of informed choice may not hold true. Patients may not be the primary decision makers, their choices may be limited, and their informational needs are likely to be idiosyncratic, depending on their individual circumstances. They may have neither the time nor the motivation to weigh options.

Consumers choose the physicians with whom they have an ongoing relationship, and they look for health plans that include those doctors in their network. But their doctors are most often the ones who decide which specialists their patients should use; specialists determine which hospitals patients will be treated at; and hospitals make the arrangements for patients' post-acute home health or skilled nursing care.

Developing social marketing strategies for publicly reported information on quality and efficiency to influence provider choice thus requires recognizing how health care decisions are made, in practice, and the central role that doctors, family members, and hospital personnel play in the decision process.

DECISION MAKING IN PRACTICE

- Consumers may or may not be the primary decision makers.
- Choices are often limited.
- Circumstances may or may not leave time for consumers to weigh options.
- Informational needs are highly specific to individual circumstances.

III. Professionals' vs. Consumers' Understanding of Quality and Efficiency

In real life, consumers look for information relevant to their personal circumstances, when making decisions about health plans or providers: Does this health plan cover the specific services I need? Does this doctor know how to treat my condition? Does this provider accept my insurance? Which hospital is my doctor affiliated with? Does the doctor or staff speak my language? Is there a nursing home close by that has a bed for my mother? If this is the kind of information consumers are looking for, how, if at all, do questions about quality or efficiency enter into the equation, and what do consumers make of the kind of information presented on the *Compare* sites?

In this chapter, we look at the differences between professionals' and consumers' perceptions and the practical implications of these differences. We begin by reviewing the professional thinking that has shaped public reporting on the *Compare* sites (and elsewhere), and consumers' take on reported quality measures. We then look more closely at how consumers themselves understand quality and efficiency in health care, and the factors that shape their perceptions. Finally, we review formative research into the development of a common framework that can bridge the gap to engage consumers in thinking about health care quality and efficiency at a broader level.

MEASURING AND REPORTING ON QUALITY AND EFFICIENCY: THE PROFESSIONAL PERSPECTIVE

Although the *Compare* sites are intended for consumers, the important decisions about what to measure, how to measure it, and how to report it are made by quality professionals and health care providers operating within a professional framework of quality oversight or quality improvement. This orientation entails a focus on the identifiable *problems* in health care delivery — standards of care that are not uniformly adhered to, areas

of performance where there are noteworthy variations. Among professionals operating in this context, a fair degree of consensus has evolved about the criteria that quality measures should meet before being adopted for large-scale monitoring or reporting initiatives: they should address important aspects of care that health care providers can control, be firmly grounded in evidence-based standards of care, and be capable of detecting meaningful differences in performance using data that are readily available.⁵

This professional quality perspective has, in turn, shaped what is reported on the *Compare* sites — for example, process-of-care measures that reflect unevenly implemented clinical standards of care; adverse outcomes that may, in the aggregate, signal underlying quality or safety problems; resource use measures that may indicate overuse or misuse of costly resources. This perspective also shapes how measures are displayed — highlighting, for example, variations in performance, statistically significant differences from national benchmarks, an emphasis on technical specificity and accuracy in measure labels, descriptions, and displays. The implicit objective is to hold health care providers publicly accountable by drawing attention to differences in performance that professionals recognize to be both clinically important and potentially problematic. Indeed, measures on which performance is uniformly strong (“topped-out” measures) have been retired from public reporting because they no longer discern differences in performance among providers.

⁵ For a summary of criteria used by the oversight organizations such as the Agency for Healthcare Research and Quality (AHRQ), the National Quality Forum, and The Joint Commissions, see “Appendix A: Measure Evaluation Framework: Selecting Quality and Resource Use Measures: A Decision Guide for Community Quality Collaboratives,” Rockville, MD: AHRQ, May 2010, available at [<http://www.abrq.gov/professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/perfmeasappa.html>].

PROFESSIONAL CRITERIA FOR SELECTING QUALITY PERFORMANCE MEASURES

IMPORTANCE

- Capturing an important aspect of quality amenable to provider or health system control.

SCIENTIFIC ACCEPTABILITY

- Having a valid, reliable, and well-defined evidence-based relationship to health care quality.

USABILITY

- Capable of discerning statistically and clinically meaningful differences in performance.

FEASIBILITY

- Measurable using readily available data that are auditable and can be collected without too much burden.

Source: Agency for Healthcare Research and Quality

Consumers do not know what to make of the quality measures that are reported or understand the rationale for reporting them. They are baffled by process of care measures, because they assume that doctors — not administrative entities like health plans or home health agencies, or even facilities like hospitals or dialysis centers — are responsible for deciding what treatments their patients do or do not need. When they think about the outcomes of care, they prefer to think in terms of sick people being cured or getting better. They know, too, that bad things can happen (like deaths or complications), but they do not understand the rationale for pointing these out on the *Compare* sites or for holding health care providers responsible for things that may be unavoidable.

Consumers are largely unaware of the problems that quality initiatives are intended to address.

Policy makers and professionals may think in terms of quality, efficiency, and value in the health care system writ large, but consumers think first as patients or family caregivers. They do not understand measures that are intended to draw attention to unnecessary or overused technologies, because they assume physicians will use their best judgment to decide what treatments patients need in any given circumstance. When consumers think about the costs of health care, they think in terms of what they are charged or what they have to pay out of pocket. They understand little about how third-party payers (including Medicare) reimburse health care providers, and they often assume that higher payment rates reflect care that is technically more sophisticated or of higher quality.

These assumptions and expectations on the part of visitors to the *Compare* sites who are not familiar with quality professionals' way of thinking pave the way for a great deal of misunderstanding and misinterpretation of publicly reported measures (as we discuss in the next chapter).

CONSUMERS' TAKE ON PUBLICLY REPORTED MEASURES

Consumers who visit the *Compare* sites have little or no context for understanding the professional perspective on quality measurement and reporting. They are largely unaware of the underlying problems that quality initiatives are intended to address, unless there has been some prior publicity around particular quality issues. They trust their own doctors to make the right decisions about their care, and they assume that Medicare certification, Joint Commission accreditation, or professional credentialing ensures that other health care professionals and institutions meet basic standards of quality and competency. What they expect to find when they come to the *Compare* sites is information about services provided (or covered), staffing, credentialing, areas of specialization or expertise.

HOW CONSUMERS THINK ABOUT QUALITY

How, then, do consumers think about quality in health care? Although there has been some effort to bring them to the quality measurement and reporting table (for example, through the National Quality Forum's institutionalized consumer/purchaser councils), their involvement has almost always come downstream in the process, with an emphasis on initiating them into the

professional way of thinking. To date, the primary quality measures that have been developed explicitly from a consumer perspective are the Consumer Assessment of Healthcare Providers and Systems family of survey-based patient experience measures — and even these have generated debate among professionals as to whether they should count as quality measures. Indeed, many quality professionals doubt that consumers can judge quality, given their focus on their own personal circumstances and experiences. The common assumption is that what consumers care most about, and pay most attention to, are the interpersonal aspects of care, or what is often referred to as “bedside manner.”

When consumers are asked what comes to mind when they think of good or bad quality health care, they do speak of doctors or health care professionals who are caring, friendly, understanding, and “treat you like a person.” However, we have found consumers to have a more nuanced understanding of quality than conventional wisdom would imply — an understanding firmly grounded in an awareness of quality in the technical sense, even if it is not in the same sense that professionals see it. What they call attention to, as described below, is a broad array of technical, structural, behavioral, managerial, and systemic factors affecting quality that they can discern firsthand.

Technical Resources and Expertise

Consumers associate good quality with doctors' and staff's qualifications, credentials, training, and experience; their specialized areas of clinical expertise; access to up-to-date diagnostic and therapeutic technologies; preventive care; availability of research-based treatments; and good outcomes. They see inexperience or lack of skills and training as a factor contributing to bad quality, including misdiagnosis, medical errors or malpractice, health care-associated infections, and bad outcomes.

Paying Attention to the Patient

What consumers emphasize most about interpersonal aspects of care is the importance of paying close attention to patients and “taking them seriously.” They see this as an important part of clinical quality — paying attention to the unique circumstances and symptoms of individual patients, for example (instead of making a priori assumptions about them); being watchful and vigilant of patients, especially when they are incapacitated; taking the time needed to attend to patients' personal care and treatment. Lack of attention or “rushing through” is also seen as a contributing factor to poor quality, leading to missed diagnoses, errors in judgment, and accidents, and leaving patients and caregivers with a sense of vulnerability.

Consumers have a nuanced understanding of quality, even if not in the same sense that professionals see it.

Management of Health Care Facilities

Consumers recognize that the managerial attributes of health care facilities can contribute to discernible variations in quality — including staffing quality and staffing levels; staff morale and burnout; the quality of supervision, training, and teamwork; operating systems that make things run smoothly; and safety and emergency response procedures.

Communication

Communication is closely linked, in consumers' minds, to clinical quality and personal attention to patients. They value honesty and stress the importance of explaining things clearly and completely to patients and family members, noting that lack of clear communication can lead to serious problems. Equally or even more important to the quality of care, in the minds of many, is the extent to which doctors or other caregivers listen to patients and pay attention to what they have to say. Consumers also emphasize the importance of communication among providers, both in terms of coordinating care and delivering a consistent set of messages to patients. Many consumers are also beginning to notice that information technology (including electronic health records and email communication with providers) is improving communication.

Physical Attributes of a Health Care Facility

Consumers often mention the physical or structural characteristics of health care facilities when talking about quality. The attribute they mention most often is cleanliness, noting that unsanitary conditions can lead to infections. Other characteristics associated with good quality include affiliation with an academic research institution, the availability of up-to-date technological innovations, and the newness of the facility. Family members, in particular, may also take notice of conveniences and amenities, such as a hospital's food or internet service.

Health System Issues

Consumers also raise issues related to the broader workings of the health care system. They often notice, for example, the lack of collaboration among specialists or providers who work in different health care settings, the efficiency (or, more often, the perceived inefficiency) of the referral and prior-approval process in some health plans, and inequities in how well the system works for different groups of people.

Operational Efficiency

Consumers value efficiency, in the sense of not “wasting my time,” and discussions about quality prompt frequent mention of wait times and timely access to care. Consumers appreciate offices or clinics that run smoothly and get patients in and out quickly, electronic access to providers for referrals or prescriptions, 24-hour nurse call-in lines, and “fast service.” At the same time, they are concerned that providers may sacrifice thoroughness, for the sake of convenience, speed, and efficiency. They do not want to be “rushed through” if it means not getting the attention they need.

Costs

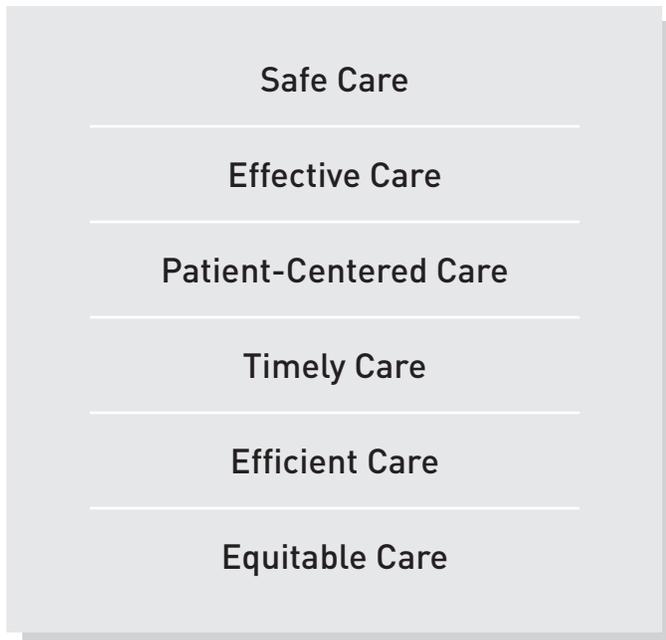
The topic of cost can also arise in discussions about health care quality. Consumers often refer to their out-of-pocket costs, in terms of co-pays and deductibles, and sometimes note high line-item charges on hospital bills. A recurring concern is that what happens in health care is “all about money.” In this context, high-quality providers are seen as those who look out for patients’ interests and provide options, instead of looking only to their own financial gain.

DEFINING A COMMON FRAMEWORK FOR COMMUNICATING ABOUT QUALITY

In policy circles, dialogue around quality and efficiency in health care has begun to coalesce around two related sets of strategic objectives reflecting some degree of professional consensus about systemic problems and areas in need of improvement: (1) the Institute of Medicine’s (IOM’s) six aims for improvement,⁶ and (2) the six quality domains defined as part of the National Quality Strategy.⁷ Other researchers’ findings have suggested that at least some of these concepts, when explained in consumer-friendly language, help enhance consumers’ ability to conceptualize quality problems or interpret comparative displays of performance on selected measures.⁸ Our research has also explored the potential usefulness of the IOM and National Quality Strategy domains as both a conceptual framework for engaging consumers in questions about quality, efficiency, and value in health care and an organizational framework for publicly reporting expanded measures of quality, efficiency, and value on the *Compare* sites.

As noted, we have found consumers’ subjective perceptions of quality to be based on concrete phenomena they can observe or experience firsthand, rather than on abstract value propositions such as those embodied in the IOM and National Quality Strategy domains. Nevertheless, they quite readily connect their firsthand perceptions and experiences to the underlying technical/structural, behavioral, managerial, and systemic factors that lead to quality within these domains. They understand, for example, that *effective* care, as the IOM defines it, is a function of the quality components they identify — including technical skills and experience of the professional staff, staff attentiveness to individual patients, the quality of the operating systems and procedures, communication among professionals and between clinicians and patients, and the availability of up-to-date technologies and research-based treatments. Although consumers may be less likely to buy into the concept of *efficiency* in the sense of “avoiding waste,” for fear that it implies restricting access

INSTITUTE OF MEDICINE’S SIX AIMS



6 Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington, DC: National Academy Press, 2001.
 7 U.S. Department of Health & Human Services. “Report to Congress. National Strategy for Quality Improvement in Health Care,” Washington, DC, March 2011.
 8 See, for example, Hibbard, J. H., J. Greene, and D. Daniel, “What Is Quality Anyway? Performance Reports that Clearly Communicate to Consumers the Meaning of Quality of Care,” *Medical Care Research and Review*, vol. 67, 2010, pp. 275–293; and Ginsburg, Marjorie, and Kathy Glasmire, “Consumers’ Priorities for Hospital Quality Improvement and Implications for Public Reporting,” California HealthCare Foundation, April 2011.

to needed or desirable services, we also found them to be quite sensitive to issues related to managerial efficiency and excessive health care costs, both of which may provide an entrée into broader conversations about efficiency and value in health care. Consumers are also aware of systemic issues that may contribute to variations in quality and inequities in health care.

This suggests the possibility of developing a common language for communicating about quality, using the IOM or National Quality Strategy domains as a framework. Relating the content of the *Compare* sites to these quality domains can help make them a vehicle for engaging consumers in discussions that move beyond the personal decisions they might face at a particular moment, but that draw on their experience to think about health care quality in broader terms.

Consumers' perceptions of quality are based on concrete phenomena they can observe firsthand.

NATIONAL QUALITY STRATEGY PRIORITIES

Safety, or Reduction of Harm

Engaging Patients and Families

Care Coordination

Effective Prevention and Treatment

Promoting Population and Community Health and Reducing Disparities

Promoting Affordability, Efficiency, and Cost Reduction

IV. Conveying Performance Information

The immediate focus of most of our research over these past 10 years has been on very practical questions about how to convey specific information reported on the *Compare* sites to a target audience that includes lay consumers and the information intermediaries (including health care professionals) who influence their decisions. The goal is to convey the information in a way that consumers can understand it, interpret it accurately, and (hopefully) apply or use it. The challenge is to bridge the gap between the needs, expectations, understanding, and real-life decisions that consumers bring to the table and the perspectives that have shaped the way that quality professionals think about, measure, and report health care quality and efficiency.

Much attention has been given in the recent professional literature to widespread problems with health literacy and numeracy (or mathematical literacy) as barriers to consumers' ability to process and use basic health information to make appropriate health care decisions. Our research suggests, however, that the gap between professionals' and the lay public's understanding of health care quality, cost, and efficiency is not simply explained by differences in education or language and math skills (as real as those differences may be). There are basic differences in the ways that quality professionals and consumers think about and

The gap between professionals' and the lay public's understanding of publicly reported measures is not explained by differences in education or language and math skills.

CONVEYING PERFORMANCE INFORMATION

ESTABLISHING THE CONTEXT

- What underlying health care problem or issue is being addressed?
- How do reported measures bear on this problem or issue?

EXPLAINING THE METRICS

- What is being measured?
- What do the numbers represent?

PRACTICAL IMPLICATIONS

- What does this tell me about differences among plans or providers in my community?
- What does this tell me about the magnitude or importance of this problem (overall, or in my community)?
- What does this mean to me?

approach these issues that create barriers even for more-educated audiences. Augmenting the *Compare* sites with consumer-friendly graphic displays and metrics and plain language alone does not help consumers struggling to know how and when to use the information.

Educators working in the area of adult learning and literacy note that adult numeracy is related as much to the context for using quantitative information — that is, the purpose and nature of the task at hand — as to the specific mathematical knowledge or cognitive and affective problem-solving processes that may be required. The dilemma in presenting health care performance information to consumers is that the context for quality professionals and for consumers is not the same. Conveying the information in a way that consumers can understand, interpret, and use requires (1) establishing a common ground for understanding the problem or issues to be addressed, (2) making the connection between the reported metrics and the problem, and (3) presenting the data in a format that consumers can interpret accurately and apply to the problem or issue at hand.

In this chapter, we review what we have learned about the particular challenges related to language, metrics, and displays for presenting information on quality, cost, and efficiency to consumers, focusing on the *Compare* websites.

ESTABLISHING THE CONTEXT: WHAT IS THIS AND WHY IS IT IMPORTANT?

Historically, the *Compare* sites have conveyed few conceptual clues to help orient lay users to the sites' overall purpose and content. Also, they have not been designed to help consumers tailor information to their particular needs or interests. Most of the *Compare* sites begin with a search function that allows users to generate a list of plans or providers available within a geographic area, from which they can select a few to compare. But our research consistently confirms that the uninitiated user who lacks any prior knowledge of the website's purpose or content has no way of anticipating what kind of information this search will yield.

Most experienced website users do not read introductory text very closely. Instead, they navigate as quickly as possible to what they assume to be the core content, taking their cues from the organizational features and labels they see. Categories and terms that make sense to Medicare program personnel or quality professionals may convey something very different to consumers. Consumers often assume, for example, that a *patient experience*

Terms that make sense to quality professionals may convey something very different to consumers.

label will lead to personal testimonials (like those they are familiar with from websites like Angie's List); that information on *Medicare payment* will explain what Medicare does (and does not) pay for; that terms like *processes of care* or even *quality measures* refer to administrative procedures.

More recently, CMS has sought consumer input to develop more-intuitive labels and organizational cues to convey information about content more accurately. The tab structure and several of the newer labels on the *Hospital Compare* site, for example, are now suggestive of the quality concepts reflected in the National Quality Strategy domains, and labels like *survey of patients' experience, timely & effective care*, and *complications & deaths* have tested well in terms of their clarity. Most consumers also readily understand what *health inspections* and *staffing* refer to (on the *Nursing Home Compare* site). However, other terms and labels on the results pages of *Hospital Compare* and other *Compare* sites — notably, *readmissions, use of medical imaging, Medicare payment*, and *number of Medicare patients* — remain obscure and subject to misinterpretation.

Establishing the context — giving website users a general idea about what is being reported and why it is important — is critical to their understanding of the *Compare* tools, as they are currently designed. When we direct research participants' attention to narrative explanations in testing situations, they typically find this very helpful in making sense of what can otherwise seem a bewildering display of technical information. However, they rarely find or see this information on their own.

ORGANIZING AND LAYERING INFORMATION

Another major challenge consumers face in comprehending or making use of the health care performance measures on the *Compare* websites lies in the sheer volume and complexity of the information that is (and, increasingly, will be) publicly reported. In our approach to consumer research, we typically try to show research participants actual performance metrics as they would likely be presented alongside other information on the *Compare* sites, in order to get a realistic sense of how they would react to it in practice. Consumers will scan data displays, we have found,

to get a sense of what is being reported, but they will do little more than gloss over them if confronted with a lot of detail at once. Often, in testing situations, they will focus on the first two or three measures displayed but will not seem to notice or pay attention to any others.

Recognizing this problem, consumer advocates have recommended simplifying quality reports aimed at consumers through summary ratings or composite measures of performance. We, too, have found that consumers often want and expect to see such ratings, and may even wrongly assume that the first displays they see *are* overall performance ratings. (In the very first display on the *Home Health Compare* results page, for example, consumers sometimes misinterpret the symbols indicating services that are (✓) and are not (⊘) offered as overall ratings of acceptable and unacceptable performance in these areas, and look no further.) At this writing, only the *Nursing Home Compare* and *Medicare Plan Finder* websites provide summary ratings, but there has been considerable interest within CMS in developing and reporting them on the other sites, as well.

Consumers will gloss over data displays if confronted with a lot of detail at once.

However, we have also found that consumers often get the wrong idea about what summary ratings actually mean, in the absence of any prior knowledge of what goes into quality measurement and reporting. They tend to interpret such measures in light of what they expect to see, which can be misleading. Consumers typically assume, for example, that overall ratings of health care providers or facilities reflect medical experts' judgments about the comprehensiveness of the services offered, the credentials and experience of medical professionals, and a host of other characteristics that are not usually captured by the metrics.

Although consumers often expect and want to see summary measures of performance, many are also interested in drilling down to the more detailed information on the *Compare* sites. Some will ask, specifically, what a composite score or summary rating is based on. Others will note that overall ratings can obscure important variations among departments, services, or clinicians within a health care facility. Many consumers look for ways to select or tailor information to their specific interests or needs. However, the data currently available on the *Compare*

sites often do not readily fit the search categories or provide the type of information that consumers care about. Others look for search or filtering mechanisms that would allow them to sort information by medical conditions or procedures not currently included in performance metrics.

EXPLAINING WHAT IS BEING MEASURED AND HOW IT RELATES TO QUALITY

Orienting consumers to the value of publicly reported information on health care quality also requires clarifying, at the measure level, what is being measured and how it relates to quality of care. Consumers take their first cues from the category and measure labels themselves, which vary considerably from one *Compare* site to another. Confronted with displays of technical information they do not understand, website users will typically look for (and click on) any link that looks as if it will provide an explanation — for example, the measure label itself, an information icon (i), or a link to “more information.” If they do not find what they are looking for with one or two clicks, they typically become frustrated and look no further.

Findings from consumer testing of technical measures on the *Compare* sites raise questions about how best to balance clarity with brevity, and specificity with “skimmability.” Although many of the measure labels on *Hospital Compare*, for example, have been drafted to convey as precisely as possible what is being measured, this can result in wordy labels that are hard for users to skim through. Labels on the *Medicare Plan Finder* site, by contrast, tend to be short and easy to skim, but users often misinterpret them. For example, the “Breast Cancer Screening” measure, which actually reports the percentage of women who received a mammogram, is usually understood to mean how good the plan's coverage is for breast cancer screening.

Technical language and professional jargon can also hinder consumers' understanding of publicly reported measures. Terms that may have specific meaning to health researchers, clinicians, quality professionals, or Medicare staff often have a different meaning in the everyday language of consumers.

Terms that have a specific meaning to health researchers often have a different meaning in the everyday language of consumers.

Consumers react emotionally to some terms, even though their meaning is quite clear.

For example, in consumer testing of hospital readmissions measures, a number of research participants assumed that the term *rate* (as in “hospital readmission rate”) referred to the amount charged or paid for a hospital readmission, until the term was more clearly qualified (“rate of hospital readmission”). Consumers may also be unfamiliar with the use of the term *measures* to mean performance metrics, and think that *quality measures* (like *safety measures*) refer to specific administrative actions taken by health care providers. Other commonly used, but confusing, terms include *beneficiary*, when applied to a person on Medicare (confused with the beneficiary of a will or life insurance policy); *Original Medicare* (interpreted as applying to the very first Medicare legislation or people covered by it); health care *practitioners* (not understood as referring to health care professionals); and health care *providers* (confused with nonprofessional personal care providers).

Technical terms and acronyms that refer to specific medical conditions, tests, or procedures — such as urea reduction rate (URR) or Kt/V on the *Dialysis Compare* site, or venous thromboembolism (VTE) or *Clostridium difficile* on *Hospital Compare* — will also usually require some explanation. However, some acronyms (like MRSA, for methicillin-resistant *Staphylococcus aureus*) are much more familiar to consumers from news stories.

Although one of the primary aims of consumer testing is to ensure that language is clear and understandable, consumers may also react emotionally to some terms, even though their meaning is quite clear. Consumers understand “complications” and “deaths,” for example, but sometimes say they would avoid looking at this information, because it is too frightening. Death rates on the *Dialysis Facility Compare* site are reported instead as “survival rates,” but this term, too, appears to conjure up unwelcome thoughts to many consumers.

Plain language alone, however, is not sufficient to make the meaning of quality measures clear to lay users. Translating the technical terms from measure specifications into plain English yields descriptions that may be technically accurate, but still not very understandable to the average consumer. Consider, for example, the “elective delivery” measure recently introduced to the *Hospital Compare* website (see below). From a quality perspective,

the underlying rationale for publicly reporting this measure is that it addresses a common reason for hospital admission (childbirth) among the younger and healthier population and a common practice that deserves attention because it may put newborns at unnecessary risk. This, however, is not something that most lay persons are aware of. The typical approach to drafting consumer-oriented content is to start with the technical definition, substituting medical terms that may be less familiar to lay users (such as “elective” or “gestation”) with more consumer-friendly language (such as “not medically necessary” or “pregnancy”). But this, by itself, does not explain what is being measured, why, or what this has to do with quality of care.

Consumer testing revealed that several issues needed to be clarified for consumers to make sense of this measure: (1) that the measure refers to deliveries that are too early (pre-term, but not premature), induced (not spontaneous), and elective (not medically necessary); (2) that it refers to an ill-advised practice, not a benign discretionary decision on the part of patients and their doctors; and (3) that what is being measured and reported is the percentage of early deliveries that were not medically necessary (and not for example, the percentage of *all* deliveries that fell into one or both categories). These findings shaped the rather lengthy measure label and explanatory language that were subsequently reported on the *Hospital Compare* website, as shown below.

Although some publicly reported measures are more intuitive than others, the technical nature of many of them poses challenges similar to the “elective delivery” example cited above: for example, the anemia management or dialysis adequacy measures on the *Dialysis Facility Compare* site, the use of antipsychotic medications for long-term nursing home residents, or the many different blood clot prevention and treatment measures on *Hospital Compare*.

CMS has developed detailed plain-language explanations for many, if not most, of the reported measures on the *Compare* sites. When we direct research participants’ attention to them in testing situations, they often remark that they should be readily accessible to users at the outset — “*Why didn’t they say this in the first place?*” But the most immediate and visible links adjacent to the data displays do not necessarily take users to these explanations.

Plain language alone is not sufficient to make the meaning of quality measures clear to lay users.

THE CHALLENGE OF CONVEYING QUALITY INFORMATION TO CONSUMERS: "ELECTIVE DELIVERY"

TECHNICAL SPECIFICATIONS	LANGUAGE ON HOSPITAL COMPARE
<p>Measure Title</p> <ul style="list-style-type: none"> • Elective Delivery <p>Brief Description</p> <ul style="list-style-type: none"> • This measure assesses patients with elective vaginal deliveries or elective cesarean sections at ≥ 37 and < 39 weeks of gestation completed <p>Numerator</p> <ul style="list-style-type: none"> • Patients with elective deliveries for one or more of the following: <ul style="list-style-type: none"> - Medical induction of labor - Cesarean section while not in active labor or experiencing spontaneous rupture of membranes <p>Denominator</p> <ul style="list-style-type: none"> • Patients delivering newborns with ≥ 37 and < 39 weeks of gestation completed 	<p>Measure Label</p> <ul style="list-style-type: none"> • Percent of newborns whose deliveries were scheduled too early (1–3 weeks early), when a scheduled delivery was not medically necessary <p>What is this and why is it important?</p> <ul style="list-style-type: none"> • Guidelines developed by doctors and researchers say it's best to wait until the 39th completed week of pregnancy to deliver your baby because important fetal development takes place in your baby's brain and lungs during the last few weeks of pregnancy. • Sometimes women go into early labor on their own, and early deliveries can't be prevented. Sometimes, doctors decide that inducing labor or delivering a baby early by C-section (called "elective delivery") is in the best interest of the mother and the baby. In these cases, early deliveries are medically necessary. • However, doctors may also decide to induce labor or deliver babies by C-section early as a convenience to themselves or their patient. This practice is not recommended. Hospitals should work with doctors and patients to avoid early elective deliveries when they are not medically necessary. • This measure shows the percent of pregnancy women who had elective deliveries 1–3 weeks early (either vaginally or by C-section) whose early deliveries were not medically necessary. Higher numbers may indicate that hospitals aren't doing enough to discourage this unsafe practice. <p>Lower percentages are better.</p>

UNDERSTANDING THE METRICS

Consumers' interest in and ability to use the publicly reported information on the *Compare* sites will also depend on their understanding what the metrics are telling them; this, in turn, will depend on what questions they want to address. Although the emphasis among consumer advocates has been on using public reports to inform choice, "Which provider is best?" may not always be the most salient question to consumers, given their circumstances. The questions they ask about the data in testing situations suggest they also wonder, "How big a problem is this?" "How much of a risk does this pose?" "How much of a difference is there among providers?"

'Which provider is best?' may not be the most salient question. Consumers also wonder, 'How big a problem is this?' 'How much of a risk does this pose?'

How the metrics are reported can facilitate or impede consumers' understanding and use of the information to answer such questions. Quality and resource use metrics are reported on the *Compare* sites in many different ways, shaped in part by the quality professionals who are the measure stewards and other professional stakeholders. Metrics may be positively or negatively framed, with higher numbers indicating better or worse performance. They may be reported as percentages, natural frequencies, measures of time, incidence, ratios, or comparisons to a benchmark. Here, we review what we have learned about consumers' ability to interpret and use quantitative information presented in different formats.

Percentages

Many of the process-of-care and outcomes measures on the *Compare* sites are displayed as percentages, indicating the rate of compliance with a standard of care or recommended treatment (that is, the percentage of eligible patients who received the indicated service) or a risk-standardized outcomes estimate. Consumers are mostly familiar with percentages and are able to interpret them reasonably accurately, especially when they are displayed as whole numbers between 1 and 100. However, percentages less than 1 presented in decimal format may be harder to interpret.

Although percentages facilitate comparisons among providers or facilities, some consumers also want to see the actual number of cases or patients that they are based on, noting that real numbers give them a better sense of what lies behind the percentages and how many patients the metric applies to in a particular case.

Rates per Thousand (Natural Frequency)

Quality measures reflecting rarer events (such as complication rates) are often presented as rates per thousand, rather than percentages. However, consumers are less familiar with this format, and often assume the numbers reported reflect either the actual number of cases or a percentage. As a result, although they may be able to interpret the numbers for purposes of comparison, they often conclude that the actual frequency is much higher than it is. Moreover, both consumers and health care professionals (including physicians) are often perplexed to see rates per thousand that have been calculated based on fewer than 1,000 eligible cases, and may assume that these numbers are wrong.

Consumers have particular difficulty interpreting frequencies less than 1 in 1,000, expressed as decimals, as is the case for some health care—associated infections (HAIs) and hospital-acquired conditions (HACs) and other rare events. When looking at draft displays of HACs, for example, many research participants assumed that the decimals represented percentages (for example, assuming a frequency of 0.050/thousand = 5 percent, or even 50 percent), and many were unable to say how often such events would occur based on the rates shown. On the *Hospital Compare* website, rates for HACs and HAIs (rare occurrences) often differ only slightly between hospitals. Consumers may understand that these are rare, but they often still overestimate both the risk and the differences between the hospitals, assuming one hospital's performance is markedly better or worse than another's based on the numbers shown, when in reality there is very little difference between them.

Our research has also shown that presenting rare events as normalized rates, or variable frequencies normalized to 1 (for example, 1 out of 12,000 cases versus 0.084 per 1,000 cases) was a somewhat more familiar format to consumers (similar to gambling odds, for example) that better conveyed a sense of how often events occur. However, presenting the data based on

Showing the actual numbers of patients affected aids comprehension.

normalized rates makes it harder for respondents to compare health care providers or facilities to each other or to benchmarks, because the denominators vary. Showing the incidence or the actual numbers of patients affected, in addition to rates, also aids comprehension of this type of information.

Numerators and Denominators

The numerators and denominators used to calculate rates, whether percentages or frequencies, are a source of confusion for many consumers, which can impede their ability to understand or interpret the data accurately.

Consumers often want to know how many patients are included in a particular measure, in part because it conveys a sense of how many patients are affected by that quality indicator and in part because it tells them how many such patients a particular provider has seen. Some displays on the *Compare* sites show both the performance rate and the denominator in more detailed tabular data displays (for example, 99% of 880 patients or 99% of 385 patients). However, seeing the data displayed this way, with different denominators, confused some consumers, and led others to think they would have to calculate the statistics themselves. Consumers were better able to interpret these numbers when the denominator information was presented in text next to graphic displays (for example, inserting the words “based on 880 Medicare patients” next to a bar showing the performance rate as a percentage), rather than in data tables. In this latter case, consumers liked seeing the denominator numbers, both as an aid to interpreting the performance rates and as a proxy measure for provider experience.

Consumers also frequently misunderstand which patient populations are included in the numerator and denominator of different quality measures displayed on the *Compare* websites. Some measures include all patients, whereas others (often displayed in the same measure set) include only a subset of patients with a specific condition or receiving a certain type of surgery or treatment. For example, for the *Elective delivery* measure cited previously, only patients who had early deliveries are included in the denominator. Consumers often overlook this information, however, and assume that the rate shown is a percentage of all women who gave birth. This oversight can make it especially hard to interpret safety measures reflecting rare complications or adverse events and to determine which pose the greater risk, in terms of actual numbers of patients affected.

Comparisons to Benchmarks, Bucket Displays, and Ratios

Among quality professionals, the performance of an individual health care provider or facility on a specific measure of quality or

resource use is typically compared to a standard, or benchmark, that represents an acceptable, or “expected,” rate of performance. This serves as a basis for flagging high and low performers. How the benchmark is defined (and how high the bar is set) will depend on the statistical characteristics of the measure and how it will be used. But for purposes of public reporting, the benchmark usually represents an adjusted national performance rate.

Comparisons to a benchmark can mislead consumers as to how facilities compare to each other.

Consumers, too, like to see a benchmark to serve as a point of reference for making comparisons. They vary, however, in what type of benchmark they would like to see. Many consumers who approach the *Compare* sites wondering “what is best for me” are looking to make comparisons in the metropolitan area where they live, and have little interest in benchmarks representing national or state averages. Some also have little interest in “average” performance of any sort, and would rather see best or ideal performance as the benchmark.

Because of the potential for confusion or misinterpretation of some performance measures (especially dire outcomes measures, such as deaths and complications), some metrics on the *Compare* sites are reported in terms that convey comparisons to benchmarks rather than actual performance rates. A commonly used format is the so-called “bucket” display, which categorizes performance based on statistically significant differences from benchmarks (worse, no different, better; or lower, no different, higher). Consumers generally understand what these displays are conveying, and they can be useful for summarizing performance across many different measures in a single facility profile. However, categorizing performance based on comparisons to a benchmark may be of little use to consumers for comparative purposes.

Most consumers want more detail than either bucket displays or observed-to-expected ratios provide.

One issue has to do with how buckets are defined. Website users invariably assume that the middle category reflects “average” performance. Because the categories are based on statistical significance, however, the middle bucket actually captures both average providers and (more often) those whose numbers are too small to determine whether their performance is statistically significantly different. Their performance rates may be high or low, but the difference from the benchmark is uncertain. The “no different” label is intended to capture this ambiguity, but it is a distinction that most consumers (and other audiences) miss. The only providers or facilities that would be flagged as high or low are those whose volume is high enough to generate numbers that are statistically significant.

More important, simple comparisons to a benchmark can mislead consumers as to how facilities compare to each other. The fact that a facility is flagged as high or low in a bucket display does not mean that its performance is significantly different from that of the facilities that fall into the middle bucket. More often, they are not.

Another metric used to convey performance in comparison to benchmarks are ratios of observed-to-expected performance, such as the standardized infection ratios used to report HAIs, or the *Spending per hospital patient with Medicare* measure, both on *Hospital Compare*. These are usually reported as decimals, where 1.0 represents expected performance, numbers less than 1.0 represent lower-than-expected performance, and numbers greater than 1.0 reflect higher-than-expected performance. Consumers are not accustomed to seeing ratios presented this way and are likely to misinterpret the metrics as percentages (or, in the case of the spending measure, as dollars), unless the ratios are explained to them. Once the ratios are explained, consumers generally understand that, like the bucket displays, these ratios show comparisons to benchmarks, rather than actual performance rates. Unlike bucket displays, however, ratios better convey the magnitude of the difference and the degree of variation among providers or facilities.

Most consumers want more detail than either the bucket displays or the observed-to-expected ratios provide. They typically want to know what actual performance rates are, how many cases or people are affected (in the case of HAIs), how much difference there actually is among providers. In some cases, bucket displays are presented on the *Compare* sites as first-level summaries, and consumers can drill down to more detail showing actual performance rates. When that information is available, some of the limitations of the comparisons to benchmarks become more apparent to them.

Directionality and Negative Measures

Another challenge in presenting quality information to consumers in a way they can interpret accurately has to do with the intended directionality of the measure — whether higher rates reflect better or worse performance. Consumers understand directionality when looking at a single measure whose intended meaning is reasonably clear. In some cases, however, the intention is not clear: Are higher rates of follow-up mammograms a good or a bad thing? Do higher rates of spending indicate waste or higher quality care?

The intended directionality of a measure may not be clear: Are higher rates of follow-up mammograms a good or a bad thing? Do higher rates of spending indicate waste or higher quality care?

Consumers generally equate higher performance rates with better performance and may misinterpret data, especially when scanning displays of multiple measures. For example, when testing comprehension of nursing home quality measures, some respondents reasoned that the negative quality measures referred to strengths and capabilities of the nursing homes, rather than to preventable adverse outcomes. If the nursing home had a high percentage of residents with pressure sores, they interpreted this to mean the home must be particularly good at dealing with pressure sores, thus attracting a higher proportion of residents with this condition. Users become especially confused when looking at data displays that mix positive and negative measures. Under these circumstances, they may begin to question the intended meaning of measures that they previously understood.

Consumers equate higher performance rates with better performance.

Limitations of the Data and Statistical Concepts

Most consumers (and clinicians) lack a detailed knowledge of statistics and are unfamiliar with methods used for quantifying quality and performance data. For the most part, they do not have an interest in this information, and those who do can usually access descriptions of methodology through links on the *Compare* website. Nevertheless, many consumers and clinicians raise questions about the data that may affect their response to the information.

Two questions that can affect users' initial response to the *Compare* tools and their perceived usefulness have to do with where the data come from and to whom they apply. Because the tools are hosted on the Medicare website, most consumers assume that the information applies only to Medicare beneficiaries. Consumers are more inclined to trust data that have been gathered independently, but they often assume that the information is self-reported by health care providers, or is even promotional (much like information on proprietary provider websites). If patient survey data are displayed the most prominently or viewed first, users may also assume that all of the measures are based on patient surveys. In practice, the sources of information and sample populations vary across measure sets and *Compare* sites, but this information is usually provided in the more technical data details for professionals and is not readily accessible on the consumer-facing pages.

[Saying that] data have been adjusted to account for differences in risk can convey the impression that statistics have been fiddled with to make them look better.

Consumers often have an intuitive sense of some statistical concepts and potential limitations of the data, even if they are not statistically savvy. For example, most understand that a health care provider's or facility's performance on outcomes measures (deaths, complications, hospital admissions, or readmissions) will look worse if that provider treats sicker or needier patients — and they will often bring this up on their own, for example when looking at data for well-known “safety net” hospitals in their own area. Many consumers also understand that observations based on a small sample of patients will yield murky estimates (which partly accounts for their interest in knowing how many patients a measure applies to). Most consumers also recognize that small differences in reported performance rates, even if they are statistically significant, may not reflect meaningful differences

Consumers assume that missing data reflect badly on a health care facility or a provider.

in performance (which partly accounts for their interest in seeing actual performance rates, and not simply comparisons to benchmarks).

However, consumers usually do not understand statistical terms like *risk adjustment*, *interval estimates*, and *statistical significance*. Nor do they, along with many clinicians, understand (or necessarily believe) the methods of adjusting for these limitations. However, efforts to explain these concepts can raise more questions than they answer — in part, because words like *risk*, *expected*, *confidence*, and *significance* have very different meanings to the lay public than they do to quality professionals or statisticians. Explanations that the data have been *adjusted* to account for differences in risk or small sample sizes, although intended to enhance trust in the fairness of the reported measures, can convey the impression that statistics have been fiddled with just to make them look better.

Consumers are also skeptical of missing data, regardless of the terms used to explain their absence (*not available*, *no data are available for this measure*, *too few cases*, *data was suppressed for one or more quarters by CMS*, and so on). Consumers consistently assume that missing data reflect negatively on the facility or provider — that the facility does not keep adequate records, is trying to hide poor scores, that “too few cases” means patients are avoiding that particular provider, or that the provider does not have adequate experience. Missing data is one of the first things consumers notice in a data display, and they will often volunteer, on their own, that they would “stay away” from those providers or facilities.

DISPLAYS: WHAT DO NUMBERS MEAN?

In addition to the metrics themselves, how the data are displayed can influence consumers' interpretation and understanding of the publicly reported information. The *Compare* sites use several different formats.

At this writing, comparative performance metrics for selected providers are typically presented first in numeric tables, and users have the option (through links on the results page) to view more detailed data tables or to see graphic displays of the data (bar graphs). Both of these types of displays present the

measure-level metrics for selected providers, alongside state and national benchmarks (where available), and may include directional cues (especially for negative measures), but they provide no additional interpretive cues.

In addition, on some *Compare* sites and for some measures, the information may be provided in an interpretive or evaluative display format, where, in lieu of performance rates or other metrics, words or symbols convey information about how well a facility or provider performed. These formats may be used alone or in combination to convey summary or composite information as well as measure-level performance data.

Here, we review what we have learned about the advantages and disadvantages of each type of display format.

Numeric Tables

To many consumers, the sight of a table full of numbers will seem daunting — too much work to make sense of. However, some consumers prefer to see all comparative data displayed this way — notably, those who are most interested in the details. One advantage to tabular displays is that they allow users to compare the performance of many different plans, providers, or facilities across many different measures at one time. If the numbers are comparable in format and all go in the same direction, the information can be easy to scan. Tables can easily become too busy and hard to read, however, when that is not the case or when many different data elements are reported.

Graphic Displays

Many consumers prefer visual displays of data in graphical format, and most (but not all) are familiar with simple bar graphs similar to those used on the *Compare* sites. When considering a single quality measure, bar graphs facilitate at-a-glance comparisons among facilities or to benchmark. However, the use of bar graphs also presents several challenges.

We have found that consumers tend to focus more on the blue bars that represent selected providers and often either fail to notice the yellow bars that represent state or national benchmarks or not understand what they refer to. In testing alternative versions of nursing home displays, respondents were more likely to pay attention to and use benchmarks for purposes of comparison when the benchmark rate was shown as a solid line intersecting the bars.

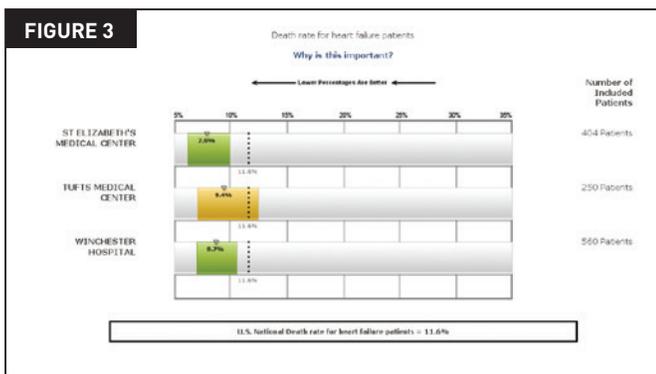
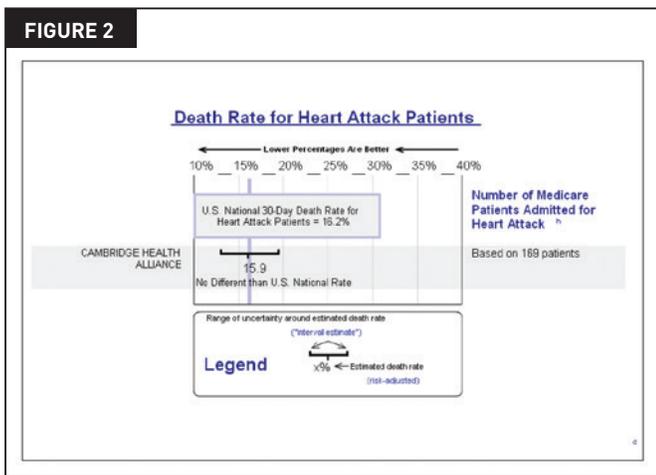
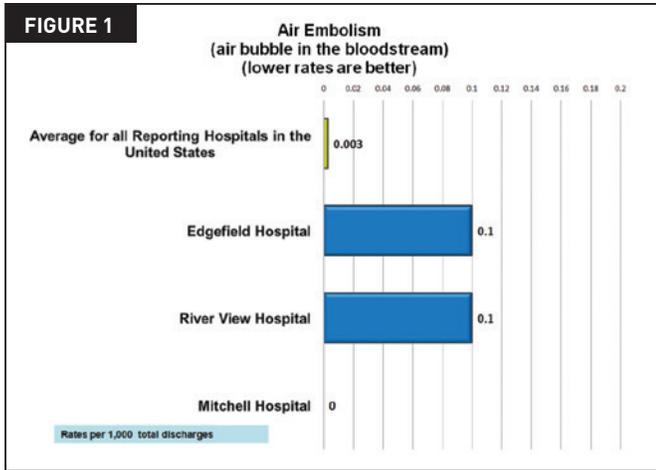
Unlike tables, bar graphs do not allow users to scan performance rates across many different measures and many different providers at one time. The *Compare* sites generate a separate graph for each

Using different scales for different measures makes comparisons misleading.

measure, with comparative bars representing up to three selected providers (in blue) and state and national benchmarks (in yellow). To view several measures, users have to scroll through several different graphs. In testing situations, they often focus only on the first graph displayed (sometimes assuming it represents an overall summary of performance) and may not notice that others are available.

Issues of scale can also make it challenging to display some information graphically. The rule of thumb is to use the same scale for all measures, preferably showing the full range of potential values (for example, a scale of 0 to 100 percent). This can be difficult when performance is measured differently for different measures or when performance rates vary markedly across measures. However, using different scales for different measures can exaggerate differences and make comparisons among measures misleading. When viewing draft displays for *Air Embolism* during consumer testing of HACs, for example, respondents at first preferred the graphs (figure 1) to the data tables because they thought they were easy to understand. When their attention was directed to the rates of occurrence, however, most respondents thought the graphs distorted the magnitude of the data.

Displaying ranges of performance in addition to point estimates can also be difficult for consumers to interpret graphically. For example, in reporting 30-day mortality measures on the *Hospital Compare* site, mortality rates are reported (and shown graphically) as point estimates along with the interval estimates. Consumer testing for the development of the first display of this measure yielded the display shown in figure 2, with the point estimate inside a bar representing the interval estimate. About half of the testing participants (including physicians) recognized that the bar represented some sort of range, although few fully understood it, and most focused on the point estimate. In redesigning the *Hospital Compare* site, the website developers changed the design to make it more visually appealing (as shown in figure 3), but without retesting to determine its impact on users' interpretation.



Evaluative Displays

Star displays, which are used on several CMS *Compare* tools, including the *Medicare Plan Finder* and *Nursing Home Compare*, assign a star rating between one and five stars based on how well a provider or plan performed. Some star ratings, like the bucket displays described above, may be based on statistically significant comparisons to benchmarks, whereas others may reflect weighted composites derived from several different measures. The stars can be used to display results both for individual measures and for composite measures and overall ratings.

Most consumers are familiar with star ratings from their use in other settings. The principal advantage of star ratings is that they allow users to skim information across many different categories or measures and across many different providers at once, and to see at a glance stronger and weaker areas of performance without having to interpret the numbers. Using star ratings also eliminates the burden of aggregating and interpreting data that may vary from measure to measure, including changes in directionality and units of measurement (percentages, seconds, number of complaints, staffing numbers). In testing situations, consumers may look at star ratings in particular categories, or count up the total number of stars to determine which provider is doing best.

However, the use of star ratings also has drawbacks. They can enhance the impression that ratings are based on expert judgment, as star ratings are commonly used in other settings (such as hotel ratings and movie or restaurant reviews), unless their statistical basis is made clear. The statistical cutoffs that determine the number of stars displayed might misrepresent the real differences in performance between providers. Also, they do not convey information about the magnitude of differences or the degree of variation.

The use of stars can enhance the impression that ratings are based on expert judgment.

Star displays may also be subject to misinterpretation when used to indicate performance on negative measures. One example of this occurred during consumer testing of star displays for a quality measure to be reported on *Physician Compare* that reflects poor blood sugar control in patients with diabetes. In this case, many consumers were confused as to whether more stars indicated better performance or higher percentages on this measure (and therefore worse performance). Star ratings also led to more errors in interpreting negative nursing home measures, when compared to the use of words (“better,” “worse”) to indicate performance. However, we have also found that this issue can be mitigated by adding an explanation to help users interpret the directionality of the measure (for example, “More stars are better because that means fewer complaints”).

Symbols other than stars, although not used on CMS’s *Compare* tools, are often used on other public reporting websites to indicate relative performance. Like star displays, symbols can be easier to skim than numeric displays, giving users a sense of overall

performance across many different categories. However, we have found that the usability of such displays depends on whether the symbols themselves are intuitive. Even commonly used symbols, like the circles used in *Consumer Reports* product reviews, may mean different things in different settings (an empty circle may indicate no points awarded, or no problems reported). When symbols are not intuitive, these displays place a high cognitive burden on users, who have to locate a legend to determine what the symbols mean and remember that information as they view the data.

TRANSLATION AND CULTURAL RELEVANCE

We have also conducted research with Spanish-speaking consumers, an audience that introduces unique considerations to the communication of information around quality and choice. In addition to the issues raised above — establishing context and explaining what is being measured and why — the additional challenges Spanish-speaking consumers face in synthesizing and using this information must be considered. For Spanish versions of the *Compare* tools, the information should go beyond mere translation, taking cultural appropriateness, norms, and relevance into consideration.

Translation alone is not sufficient to make the meaning of information clear to all users. Other factors, such as users' cultural patterns and beliefs or lack of familiarity or experience with the U.S. health system, can hinder comprehension of certain terms or concepts used on the *Compare* tools. For example, several Spanish-speaking participants were unfamiliar with certain concepts:

- Unfamiliarity with appeals or the appeals process led to difficulty comprehending measures related to appeals time lines and appeals upheld on the *Medicare Plan Finder* tool.
- Explanatory information suggesting patients should talk to a provider or ask questions about health care topics goes against cultural norms that regard doctors as experts and “authorities” who should not be questioned or challenged.

Translation alone is not sufficient to make the meaning of quality information clear to all users.

Some terms are better understood in English than in Spanish.

- The notion of “strangers coming into their home” to provide home health services raised concern for some Spanish-speaking participants, because it goes against the cultural tradition of the family providing care, until they were assured that these services would be provided with the recommendation of their doctor.

Other issues include misinterpreting phrases or terms related to quality of care, such as equating “communicate well” with providers being able to speak Spanish or “quality ratings” (“*calificaciones del plan*,” in Spanish) as denoting eligibility, or “qualifying” for a plan.

In addition, for younger or bilingual consumers, some terms are better understood, or more commonly used, in English than Spanish. For example, adult children or caregivers who are bilingual may prefer to use English language versions of written material or websites when reviewing information for their own use but may choose to view Spanish language versions of those same materials when assisting monolingual parents. In addition, several technical terms or acronyms are often more commonly used in English, even for monolingual Spanish speakers. For example, the acronym COPD (chronic obstructive pulmonary disease) is often more understood when presented in English than when presented as its Spanish equivalent EPOC; HMO is often more used than its Spanish equivalent, *organización del mantenimiento de la salud*; and technical terms such as *bronchodilators* are often more understood than their Spanish translations.

V. Usability and Navigation

Our research has also focused on evaluating the usability of the *Compare* tools, assessing users' ability to navigate, understand, and use the quality information on these tools. Usability, or user experience, focuses on having a deep understanding of users — what they need, what they value, their abilities, and also their limitations. As described above, we know consumers are most interested in information relevant to their particular tasks, needs, and circumstances. They want that information to be presented in formats that make it easy to find and use but that also provide a level of detail that makes the information useful. In addition to considering the types of information that should be included and how to display it, any website should be designed to allow users to complete the task or goal they set out to complete. Websites that have the user experience in mind support the completion of the user's task by having intentional and obvious pathways to lead them to their goal.

However, even when CMS web tool leads have the best intentions, efforts to optimize user experience can be hampered by a variety of issues. These include:

- A lack of funding to make significant revisions or improvements to the information architecture of the tools;
- Competing needs of multiple audiences — for example, the team has found that the *Medicare Plan Finder* has been optimized for information intermediaries resulting in a more difficult user experience for beneficiaries and their caregivers (this is discussed in more detail later in this section);

Those who oversee policy for the Compare tools may have different objectives than website users, and this mismatch can lead to poor user experience.

- Secondary goals of the tools sometimes impede their primary goals — for example, in addition to helping consumers select a provider or plan, the *Compare* tools aim to provide “transparency” about provider/plan performance to a wider audience of stakeholders, researchers, and the public. This latter aim is best accomplished by the inclusion of more and more information on the tools even when additional information can make it difficult for consumers to effectively and efficiently use the tool for provider/plan choice.

When the research team was asked by CMS to evaluate the *Compare* tools, we used the following key principles in our assessment of best practices in optimizing user experience. Items related to usability and navigation include the following:

Context

- Make the purpose and usefulness of the site immediately obvious
- Give users a framework for understanding what is being measured

Presentation and Display

- Use language that consumers can readily understand
- Make it easy for users to find the information they want
- Include meaningful link labels
- Use well-designed headings to help users scan and read written material

Ease of Use/Decision Support

- Minimize cognitive burden (help consumers process and synthesize information)
- Help users integrate information to make a decision
- Tailor information for different consumer audiences and the tasks they seek to undertake

Although we have not conducted recent usability testing of the *Compare* tools in their present form (the tools have changed over the years with the addition of new content and implementation of a consistent design), we have identified some overall themes through our research related to usability navigation that remain relevant.

CONTEXT

CMS periodically makes enhancements or changes to the *Compare* tools. For example, in June 2012, CMS redesigned *Hospital Compare* to add new measures and organize the information using a tab structure. Although the redesign of the *Compare* tools has simplified navigation by requiring fewer steps to access information on quality and performance measures, it eliminated some contextual information that existed in earlier versions that provided users with a *clear sense of the site's purpose and content*. These include navigational cues required by users to understand the tools' full potential. As a result, users remain unaware of the breadth of information available in the tools or that they are designed to enable users to conduct a side-by-side comparison of hospitals, home health agencies, or health plans.

The information that currently appears above the fold on the home page of each tool, including the heading "Find a [insert provider type]" directly above the search box, effectively conveys that the *Compare* tools can be used to search for providers or plans by location. Our research indicated that most users approach the tools as a way to generate a list of available hospitals/agencies/plans, often stopping upon reaching the initial Results page. For example, when evaluating the usability of the sites, we found that few users were aware that quality data existed in the *Medicare Plan Finder* and *Home Health Compare* tools, and most users did not access the quality data or the Compare page on their own. Instead, consumers and professionals thought the data available on the Results page — such as provider name, location, and basic provider characteristics — represented the full extent of available data on the tool. They also thought the "compare" feature was represented in the way the information was displayed on the Results page, because they could view the same type of information for multiple providers at one time. When prompted, users found the functionality to compare up to three providers (for example, hospitals or agencies) very useful in helping to constrain the often large number of available options.

In addition to providing users with information on the purpose and the information available on the *Compare* tools, it is helpful to give users an *overarching framework to understand what to expect*

and how the information on the sites is organized. We conducted formative research around the potential use of the IOM's six aims for improvement and the six quality domains of the National Quality Strategy as a potential conceptual or organizational framework for reporting hospital quality and efficiency measures on *Hospital Compare*. Participants understood the underlying concepts and found them to be a useful way to conceptualize *quality* in the health care setting.

During this same research, the team introduced *Hospital Compare* to participants and briefly explored their response to it, in the context of discussions earlier in that research around health care quality. Because the website had just been redesigned to include a tab-based navigation design, rather than quality measures listed as either *process of care* or *outcomes of care* measures, the team was able to probe on whether the recently revised format and tab structure facilitated users' understanding of the information on the site. Earlier research consistently found that several features of the old *Hospital Compare* format made the website difficult for consumers to navigate and understand.

Most users approach the tools as a way to generate a list of providers, stopping after the initial Results page.

Overall, participants understood the new tab structure and organization of the site, identified areas of interest, and seemed less overwhelmed by the volume of information than was the case in prior consumer research. They also understood that three of the tab labels — Patient Survey Results, Timely and Effective Care, and Readmissions, Complications, and Deaths — related to the IOM quality domains previously discussed, which were also the three topic areas of most interest to them. Although, as in prior research, participants were interested in Patient Survey Results, most were at least equally interested in the other two categories. When asked to compare these three tab labels to the corresponding IOM domain labels, most thought that labels on the website better captured the kind of information that was being conveyed.

Consistent with previous findings, however, participants had difficulty interpreting the remaining three tab labels — Use of Medical Imaging, Medicare Payment, and Number of Medicare Patients. Most participants assumed that the Use of Medical Imaging tab would show information about the availability of specific imaging services; however, when they looked at the measures under this tab more closely, a few participants recognized that they reflected potential overuse of medical

imaging, consistent with the IOM concept of avoiding waste. Participants expressed little interest in the Medicare Payment and Number of Medicare Patients tabs, which they did not associate with questions about quality.

These findings, although specific to *Hospital Compare*, indicate that CMS has made great strides in improving the design of its *Compare* tools and thus providing users with a better framework for understanding what information is included.

PRESENTATION AND DISPLAY

Overall, the *Compare* tools do a good job of providing information in plain language. Some of the tools, such as *Hospital Compare* and *Medicare Plan Finder*, organize the information in clearly labeled categories (rather than lengthy lists of individual measures) to help users navigate the site. However, some of the tools include pages that are content heavy and/or require a lot of scrolling; link labels could be used more effectively to aide navigation through the tools. For example, although the quality measures on *Home Health Compare* are grouped into five overarching categories, such as “Managing Daily Activities” and “Preventing Harm,” the default display is to expand each category and have all of the measures visible. This requires the user to scroll down the page to determine available information and identify pertinent content, rather than detect relevant information at a glance.

Additionally, the *Compare* sites sometimes provide key information below the fold. Many users are unaware they need to scroll down the page, and they often do not access this important information. This issue is most problematic when viewing the Results page and Compare page for each tool. Elderly website users may typically use a lower screen resolution for larger screen images, which would further limit the information seen above the fold.

EASE OF USE/DECISION SUPPORT

The *Compare* tools are effective resources in determining which hospitals, home health agencies, dialysis facilities, or health plans are available in or serve a given area, as well as in identifying certain features and quality data for a particular provider or plan. All of the tools use some visual displays (graphic displays, star ratings, pop-up explanations), guidance on directionality, and benchmarks to help users make sense of individual measures and discrete data. For example, *Medicare Plan Finder* uses icons on the Results page that denote high and low performance plans, enabling users to identify these plans at a glance. The tools also

provide links to explanatory text that is useful to understanding and correctly interpreting the quality measures (for example, “What is this?” or “Why is this important?” explanations) from multiple points, improving user access to this information. However, for the most part, the *Compare* tools are more optimized for counselors and intermediaries than for consumers. Several areas in which the tools could be improved to support provider/plan choice decisions are described below.

Tools tend to lack effective methods for narrowing results. For some tools, such as *Home Health Compare* and *Hospital Compare*, the initial results pages do not include the types of information, such as summary ratings, that would facilitate narrowing options in a meaningful way; instead, users are only able to limit results by physical address, provider name, and services offered/hospital type. In many instances, users are unable to limit their results to a manageable number, despite using the available filters. For example, an initial search for home health agencies in Chicago, Illinois, yields 669 agencies. Using the “Services Provided” filter to keep only agencies that offer all services eliminates 59 agencies, leaving 610 results. A user would then need to randomly select three providers at a time from among the 610 to access information, such as quality data, that could help them select a provider. The inability to narrow based on information such as quality impedes the effective use of quality information in the decision-making process.

The inability to narrow a search impedes the effective use of quality information for decision-making.

Medicare Plan Finder does include the types of “meaningful” information on the results page that is lacking from some of the other tools; however, the filters themselves hinder users’ ability to effectively narrow their results. The most significant issue is that the filters appear on the page prior to the initial results page, and users often either overlook the filters or do not feel ready to use them before they have seen their results. When users proceed to the initial results page, there are no filters available to them, and we have rarely seen a research participant return to the previous step to use the filters. Additionally, when asked to use the filter feature, most participants did not click on the “Update Results” button, which is required for some of the filters to take effect. For example, *Medicare Plan Finder* users did not realize that results were not immediately updated upon selecting a value, and they did not realize they had to click on the “Update Results” button for the filter to work. Instead, they would adjust the filter and then click on the “Continue to Plan Results” button. Clicking on

this button sometimes updated the plan results and sometimes did not. Several *Home Health Compare* users clicked on the “New Search” button instead of “Update Results,” which brought them back to the *Home Health Compare* home page. They explained that they were unfamiliar with the “Update Results” terminology and did not know what would happen if they clicked on that button.

Information is not filtered or focused by the specific tasks tool users want to accomplish. During our research, most users would initially state they were interested in viewing the quality ratings for the plans/agencies/hospitals. Upon closer examination, though, the typical user became overwhelmed by the volume of information and was not sure what to make of it or how to use it. This was particularly apparent during testing of *Home Health Compare*, which does not offer a summary or composite rating for agencies. Users struggled to develop an overall assessment from the individual quality measures as well as determine how two or more agencies compared when one agency scored better on some measures and worse on others.

The *Medicare Plan Finder* results include a summary rating of quality, but also include many different types of cost and benefits information, which poses several challenges to consumers, including the following:

- Because of space constraints, the labels for the information are generally not detailed enough for consumer audiences to fully understand the meaning of the information.
- Each cell in the results page includes several types of information, making it difficult for users to skim the page to compare specific elements.
- There are too many different types of information for a consumer to integrate to effectively narrow their results.

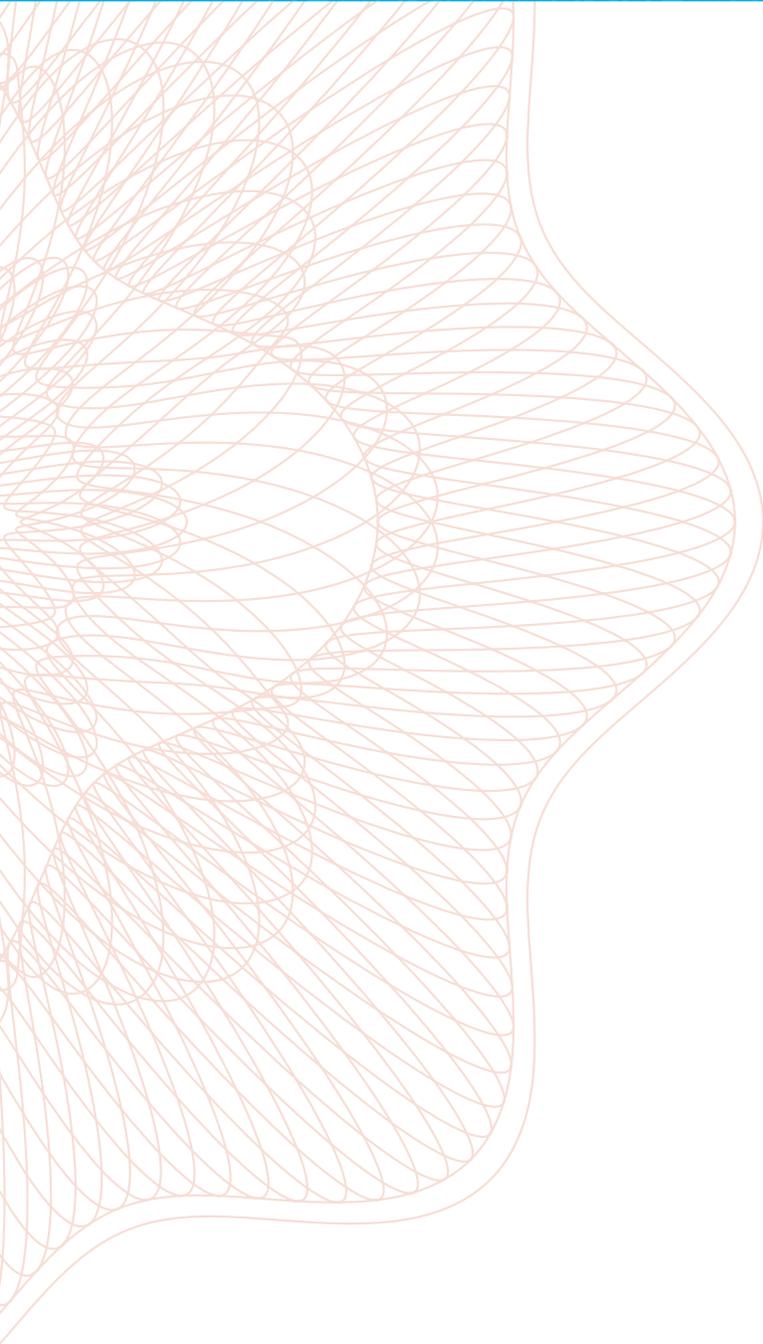
Tool design is not always optimized for consumer audiences. All of the *Compare* tools have multiple audiences, including consumers, intermediaries, and providers, and it is simply not possible to optimize them for every possible type of user. If CMS considers the primary audience for the tools to be consumers, or if the agency would like to increase and improve consumer use, then the tools should be optimized for the consumer user.

The results page of the *Medicare Plan Finder*, which we described above, provides a good illustration of this issue. Because intermediaries, such as State Health Insurance Assistance Program counselors and 1-800-MEDICARE customer service representatives, find it helpful to minimize “clicks” and see more information on a single page, the format and comprehensiveness of the initial results page is well-suited to them. Not surprisingly,

our research with intermediaries indicates that they tend to be satisfied with the tool. Given that information intermediaries help close to 100,000 beneficiaries enroll in Medicare coverage on the *Medicare Plan Finder* each year, a tool that successfully supports this audience is a notable accomplishment. However, it comes at a cost to consumer users, as the addition of more information to support intermediary use contributes to a results page that is difficult for consumers to use. A more consumer-friendly tool would limit information on the initial results — providing the elements CMS deems most important for narrowing plans to a manageable number — and then would make the more detailed information available in later pages, such as the side-by-side comparison or plan profile. Such an approach would help consumers narrow plans and prioritize information, while still making the detailed information available.

Tools could provide more-customized results. Because users access the *Compare* tools for a variety of tasks and reasons, CMS may want to explore expanding the tools’ decision-support capabilities to enable tailored sessions based on who the user is and/or the purpose of his/her search. Allowing users to input what they are looking for at the beginning of a session will likely increase the perceived utility of the tools. Potential data sources that could provide insight on the range of user goals could include data collected from the 1-800-MEDICARE call center, State Health Insurance Assistance Programs, ForeSee Surveys or Google Analytics.

Enabling tailored sessions based on users’ specific needs can expand the Compare tools’ decision-support capabilities.



www.LMpolicyresearch.com

www.mathematica-mpr.com

1743 Connecticut Avenue NW Suite 200 • Washington DC 20009

