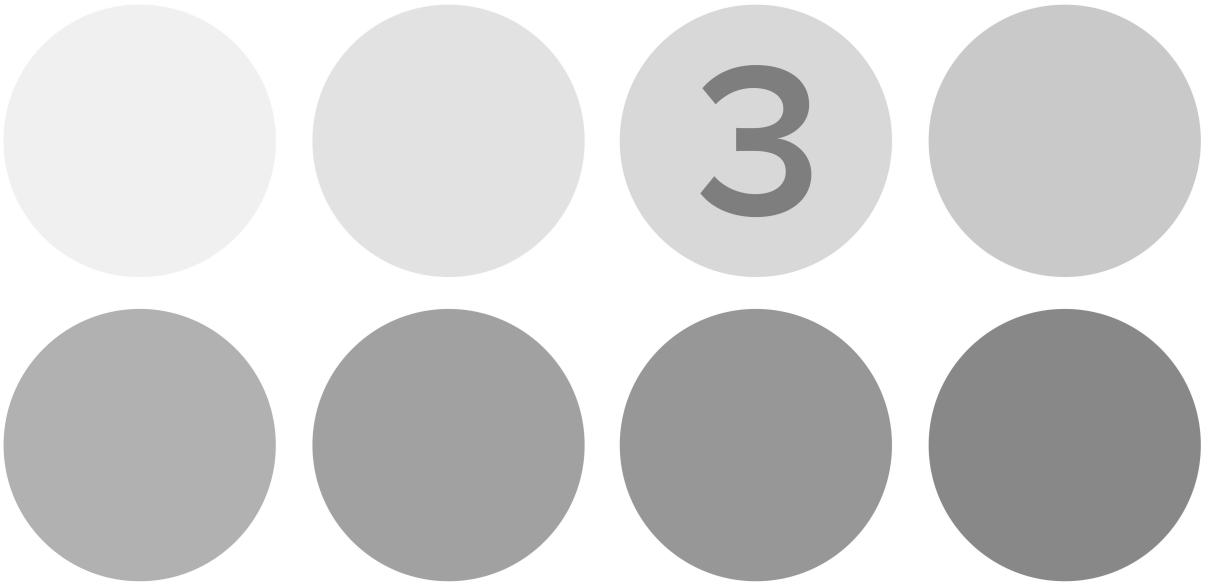
A Guide to Health Reform

Eight Practical Steps

Step Three:

Assessing Health System Performance



HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH
Department of Global Health and Population

By Michael R. Reich,
Paola Abril Campos,
Anuska Kalita,
Anya Levy Guyer,
and Winnie Yip

Step Three: Assessing Health System Performance and Defining Performance Problems

Step Three is systematically assessing the health system in order to identify problems to address through reform. This involves measuring various dimensions of health system performance, and then using the findings to help you select specific performance problems for deeper examination and reform interventions. The primary objective of Step Three is to identify both the areas in which your health system is performing well and those in which it performs poorly. This information allows you to select which areas your reform will focus on; it also establishes a baseline for monitoring and evaluating changes over time (as described later in Step Eight). Step Three has four key actions:

Key actions in Step Three:

☑	Decide what to assess, including the kinds of performance problems to assess (based on the intermediate and final performance objectives), and the types of analysis and analytic skills required
☑	Decide who will do the assessment, considering both external analysts outside of government and people internal to the Health Reform Team and government agencies
☑	Design the assessment, including the scope of assessment, time and resources required, existing data sources, and new data to be collected, with deadlines for deliverables
☑	Analyze both primary and secondary data to generate a comprehensive assessment, identify major problems, and prepare for diagnosis (Step Four)

Decide what to assess

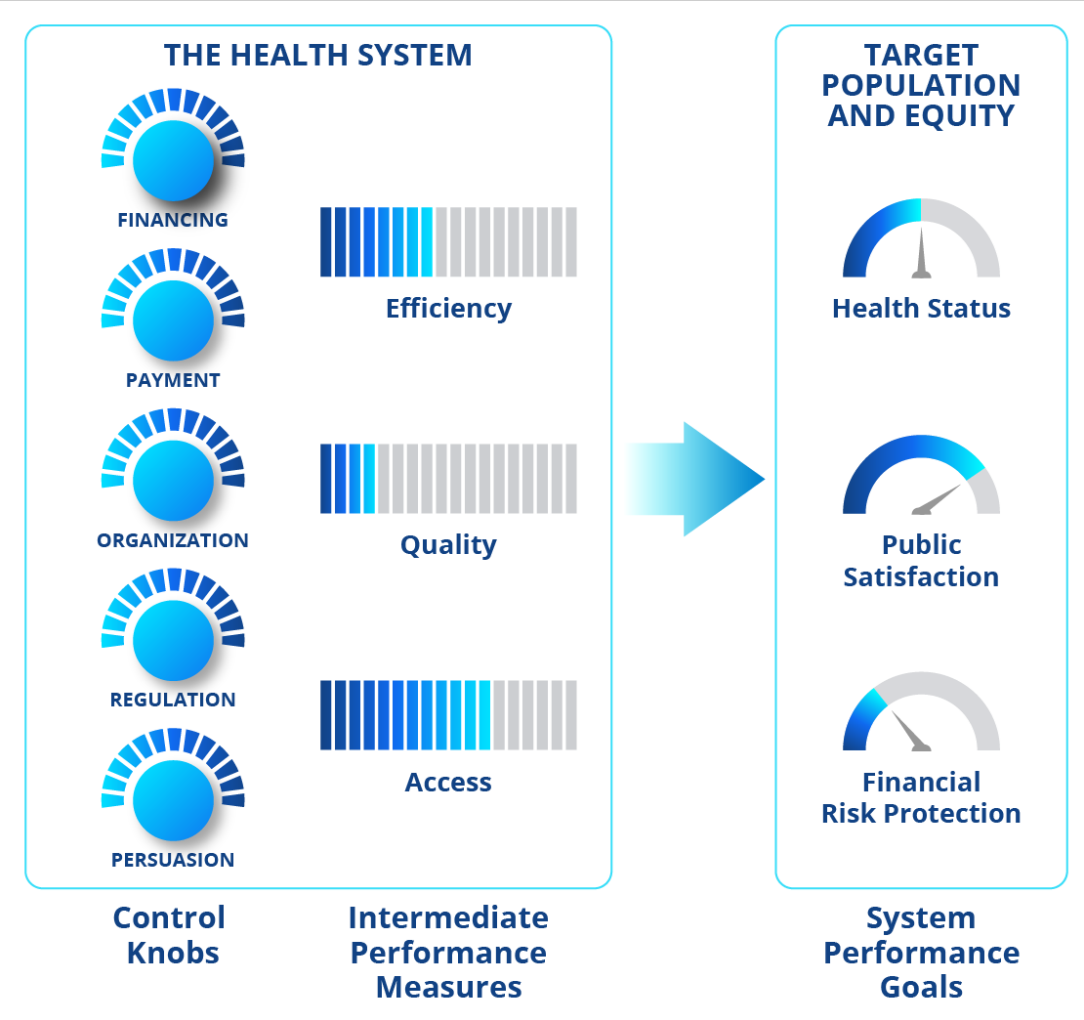
The first action is to decide which aspects of health system performance you will assess. This requires a clear understanding of how to define and measure the "performance" of a health system.

What is a health system? As noted in the Introduction, both *GHRR* and this Guide view the health system as a means to achieve certain ends. In our model, the means cover five policy controls (also called control knobs, or policy levers) that are commonly available to policymakers: financing, payment, organization, regulation, and persuasion. The ends are represented by three final outcomes for a health system and three intermediate outcomes. The three final performance objectives are: (1) health status, (2) financial risk protection, and (3) public satisfaction with the health system (*GHRR-Chapter 5*). The three intermediate outcomes (which can be influenced by the policy controls) are: (4) access to health services, (5) quality of care, and (6) efficiency in the health system (*GHRR-Chapter 6*). In our view, how well a health system performs should be assessed by how well it achieves these six objectives. These achievements, according to this model, should be assessed in two ways: the level of achievement, and how equitably the benefits achieved are distributed across the target populations. (There are, of course, other models of health systems and health system performance—these are not addressed in this Guide.)

Figure 3-1 shows the *GHRR* model of health system performance, illustrating how the control knobs affect the intermediate performance measures, which, in turn, determine the achievement of

the three ultimate performance outcomes. Appendix 3-1 presents detailed definitions for the six performance outcomes and outlines common ways to measure and interpret each one.

Figure 3-1: Health system control knobs, intermediate and final performance measures



Source: adapted from GHRR, p. 27.

What you choose to assess, and how you conduct the assessments, is determined by several factors, including: how much time is available to generate and analyze evidence, the timelines of key decision-making processes, the priorities of supportive political leaders, the available analytical capacity, and the budget. The spectrum of assessments ranges from *comprehensive* (such as the India Health Systems Research Project's assessment of the health system in the state of Odisha (Yip et al. 2022)) to *simple*. A comprehensive assessment typically involves collecting primary data, a sophisticated design and expert analyses—and therefore requires a significant amount of

time and budget. A simple assessment uses existing data and examines fewer performance measures, and is thus less resource-intensive. However, it may produce a less detailed, and less reliable or narrower assessment. Most assessments fall somewhere between the extremes.

Deciding what you will assess has important implications, as the assessment influences the scope and timeline of the health reform process. It is therefore a topic for discussion with top political leaders in addition to the Health Reform Team. Assessment decisions shape (and reflect) the overall directions of the health reform process. Deciding what to assess is not a simple technical decision; like every step in the health reform process, it has political and ethical implications.

Decide who will do the assessment

Finding the right group to undertake the assessment depends on two key factors: 1) what is being assessed, and 2) what resources are available for the assessment. Regardless of the scope of the assessment, some governments prefer to hire external assessors, such as consulting agencies or academic experts, often selected through a competitive bidding process. Conducting a comprehensive assessment of all six performance outcomes typically requires hiring an external group of experienced analysts and a substantial budget. Commissioning a comprehensive assessment, with primary data collection, of health system performance of a nation (or a state in a large country) can easily cost one million US dollars or more, involve dozens of analysts, and require two years for data collection, analysis, and report development.

The other option is using an internal assessment group (such as one located within a government agency or a government-related research group). This option has different risks and benefits. It may cost less. However, the quality of the assessment may be compromised if the group does not have the right expertise or experience. Using an internal group can also constrain the objectivity of the assessment if the analysts are subject to pressure from government officials seeking to influence the results.

The ideal health system assessment team has people with extensive experience in assessing *system performance*. Typically, the team needs members who have expertise in health system analysis, quantitative research methods (for designing and analyzing household surveys, claims data, medical records, and other large datasets), and qualitative research methods (for designing and analyzing key informant interviews, focus group discussions and other qualitative datasets). The Health Reform Team should work closely with the assessment group in order to facilitate access to data sources (for example, administrative data or policy guidelines), provide necessary financial resources, and offer overall guidance. Close collaboration and communication between the assessment group and the Health Reform Team can ensure that the health system assessment is aligned with the ethical, political, and economic priorities of the overall reform effort.

Design the assessment

The design of the performance assessment must be informed, first and foremost, by the decision about *what to assess*, i.e., which performance outcomes are the focus of the assessment. Defining the key questions for the assessment determines the appropriate methodologies to use, the relevant data sources, and the time and resources required. A well-designed assessment has data collection

tools linked to each performance outcome. Thus a comprehensive assessment likely requires a combination of existing and new data sets, and both quantitative and qualitative methodologies, while a simple assessment needs less data and fewer resources.

Identify and analyze existing data sources

Regardless of the assessment's scope, the first activity (for either the Health Reform Team or the assessment group) is to do a landscape analysis of available information and secondary data about the different performance outcomes. By identifying existing data sources, you develop a sense about the extent to which these data can inform the assessment, discover critical data gaps, and identify questions that require collection of new data.

Analyses of secondary data are an important part of the assessment's design. Health systems generate reams of statistics from different sources like management information systems, insurance claims systems, national, state, and district-level health surveys, and national and state health accounts, to name a few. The OECD and World Health Organization also collect significant amounts of national data on health. The assessment team should determine which data sets are relevant for assessing the selected performance outcomes. (Some common data sets and their interpretations are presented in Appendix 3-1 by performance outcome.) Carefully collating and analyzing secondary data can go a long way towards generating a broad assessment of the health system on several performance outcomes. Even if secondary data are not sufficient for the full assessment, they can be used to begin the analysis, identify important data gaps and inform decisions on new data collection.

The decision of how much to rely on secondary data depends on the availability and quality of the data. Most high-income countries (HICs) and some middle-income countries (MICs) like Brazil and Malaysia have extensive and robust data sets that could allow health system assessments, covering almost all six performance outcomes. However, many low- and middle-income countries (LMICs) and most low-income countries (LICs) have limited or irregular secondary data.

Additionally, not all six performance outcomes have equally complete and robust data. Most countries can effectively assess their populations' health status and access to care with secondary data from vital registrations and Demographic and Health Surveys. Most countries also have some data on public satisfaction and financial risk protection, or can add a few additional questions to existing national or state-level surveys. For the quality outcome, however, most countries will likely need to design new assessments, as many health systems do not collect clinical effectiveness or patient safety data through national surveys or health information systems.

Identify data gaps and collect new data

By doing a landscape analysis of data sources and analyzing secondary data, the Health Reform Team will likely discover several important data gaps. You may then choose to undertake (or commission) new research to allow for a more complete assessment of the health system or to develop a nuanced understanding of underlying causes behind poor performances (more on this in Step Four, Diagnosis).

Designing new research starts with defining research questions and selecting appropriate methods to answer them. Then the assessment group can design data collection instruments. Depending on what data you are seeking, these may include survey questionnaires, clinical vignettes, or interview guides. Whenever possible, we recommend utilizing indices and instruments that have already been validated and used internationally. However, if there are research areas without comparable indices, the assessment group might have to create new instruments and go through the process of validating them locally. Again, Appendix 3-1 lists some existing instruments and data sets that have been used globally to measure health system outcomes.

Data collection is usually outsourced when primary data collection is required, for two reasons. First, significant expertise in designing and collecting data is necessary to ensure quality. Second, using a third party for data collection helps to maintain independence and objectivity. The necessary skills for data collection teams, and the costs involved in data collection, will vary depending on the type of data and the size of the data set. For example, a nationally representative household survey will need a large team; in some countries, you will need multiple teams fluent in different local languages. Conducting chart reviews or standardized patient interview for assessing quality, on the other hand, needs data collectors with clinical training.

Here again, the Health Reform Team will inevitably need to consider available funds and timelines. An assessment that uses secondary data is both faster and far less expensive than one that requires collecting new data. New research could cost anywhere between a few thousand US dollars to upwards of a million US dollars, depending upon the scope and research questions. For example, a hospital chart review to assess clinical effectiveness may be fairly quick and inexpensive compared to a large household survey to assess financial risk protection or public satisfaction. Time is another important consideration. If you are trying to link your health reform effort with election or budget cycles, there may not be time for extensive primary research. In short, the Health Reform Team will decide on the assessment design based on these contextual realities and the proportions of secondary and primary research required.

A comprehensive health system assessment with both secondary data analysis and extensive primary data collection using the *GHRR* model was undertaken in the Indian state of Odisha by the Harvard India Health Systems Reform Project (Yip et al., 2022). Secondary data were used to assess health status and benchmark outcomes like financial risk protection and access to health services (outcomes where secondary data covered some but not all indicators). Health system assessments in Malaysia and Turkey using the *GHRR* model were conducted primarily using secondary data, with only limited new data (Atun et al., 2019; Johansen & Guisset 2012).

Thus, health system assessment does not always require resource-intensive collection of new data. While desirable if time and resources permit, a comprehensive assessment with extensive primary research is not necessarily a prerequisite for health reform. An assessment based on secondary data analysis, or assessing some but not all of the performance outcomes, may be sufficient, depending on your specific objectives for improving health system performance.

Analyze the data to prepare for *Step Four: Diagnosis*

The rest of Step Three is analyzing the various data collected using appropriate quantitative and qualitative analytical approaches. Once the assessment team has generated the relevant statistics and performance measures, they should work with the Health Reform Team to identify and compare your health system's performance to suitable benchmarks.

Identify benchmarks for comparison

Comparing the findings on your system's performance metrics with benchmarks is necessary to interpret whether performance on a particular outcome is good, average, or poor (see *GHRR*, Chapter 6, p. 123). For example, simply stating that "a country's infant mortality rate is 20" or that "out-of-pocket expenses constitute 30% of the country's total health expenses" does not give you a sense of *what the findings mean*.

Thus, statistics must be compared against standard benchmarks or measures. An example of a global standard measure is: out-of-pocket (OOP) expenses on health that exceed 10% of a household's total consumption are considered "catastrophic." Knowing this, you can tell that the 30% finding cited above indicates very poor performance.

Benchmarks can be determined using national averages, statistics from other states within the same country, or from other similar countries. For example, countries commonly compare health status indicators such as mortality rate and life expectancy with other countries with similar levels of economic development. For clinical effectiveness measures, clinical guidelines and standard treatment protocols are generally used as the benchmarks, as they are usually highly standardized and accepted across countries. Benchmarks for other outcomes, such as public satisfaction and patient satisfaction, are less standardized.

If improving equity among groups within your population is one of the goals of your health reform, you may choose to use internal benchmarks. Consider a health reform intended to reduce infant mortality: if the national infant mortality national rate is 20 per 100,000, but the rate is 3 per 100,000 among the wealthiest 10% of the population, you may choose to use the latter figure as the benchmark against which you compare your reform's overall performance.

The assessment group needs to identify (and provide justification for the selection of) appropriate benchmarks for comparison as part of analyzing your data. In addition to the additional details provided in Appendix 3-1, *GHRR (Chapter 6)* includes a discussion of different benchmarking strategies.

Summary

In Step Three, the Health Reform Team conducts a health system assessment to understand how well or poorly their health system performs on its intermediate and final performance goals. This will help you prioritize which performance goals to focus on for reform. Most importantly, the assessment lays a foundation for Step Four, Diagnosis, by identifying performance problems that need further examination. It also lays the foundation for Step Eight, Evaluation, by establishing baselines and identifying what should be monitored and evaluated going forward.

The *GHRR* framework emphasizes the importance of data analysis and health system assessment. However, in reality some health reform efforts have been carried out without rigorous assessments or even baseline data. This may seem like an easier path (especially when the reform effort faces time constraints or when "everyone knows" what the problems are). The reform process may even proceed smoothly. However, these efforts typically fail to generate meaningful improvements in the final performance outcomes of health systems. Without first understanding the status quo (by conducting an assessment in Step Three) and then investigating the root causes of the problems identified (which will be the focus of Step Four), you cannot confidently select appropriate reform options.

The assessment of health system performance is a foundational step in doing health reform. And it need not be a one-time activity. Health Reform Teams may decide to undertake assessments of different performance outcomes at different points in time. You may also decide to vary the scope and depth of assessments based on contextual factors, such as the availability of resources, current political priorities, policy timelines, and windows of opportunity for change. In addition, these assessments can serve as the baseline for monitoring and evaluation of the impacts of the reform, as discussed below in Step Eight.

References

Atun R, Berman P, Hsiao W, Myers E, Yap WA, et al. 2016. Contextual Analysis of the Malaysian Health System. Government of Malaysia. https://www.moh.gov.my/moh/resources/Vol_1_MHSR_Contextual_Analysis_2016.pdf

Johansen A, Guisset A. 2012. Successful Health System Reforms: The Case of Turkey. WHO Regional Office for Europe. https://disab.saglik.gov.tr/Eklenti/2106/0/successful-health-system-reforms-the-case-of-turkeypdf.pdf

Yip W, Kalita A, Bose B, Cooper J, Haakenstad A, Hsiao W, Woskie L, Reich, MR. 2022. Comprehensive assessment of health system performance in Odisha, India. Health Systems and Reform 8(1):2132366–2132366. https://doi.org/10.1080/23288604.2022.2132366

Appendix 3-1: Measuring health system performance outcomes

This appendix details the following information for the six health system performance outcomes (health status, financial risk protection, public satisfaction, access, quality, and efficiency):

- How is the outcome defined?
- How is it commonly measured?
- What data are generally available?
- What are common interpretations of this outcome (i.e., what can and cannot be concluded from the relevant data)?
- What are the common data gaps?

For further information, refer to Appendix 3-2 of this Guide (which lists additional resources for understanding assessments of the different outcomes), to *GHRR,* and to the India Health Systems Reform Project website (https://www.hsph.harvard.edu/india-health-systems/).

Final outcome #1: Health Status	
How is it defined?	The first goal of a health system is to improve the "health status" of the population it serves. Health status refers to health outcomes, indicated by disease prevalence, disease incidence, morbidity rates, and mortality rates in the population or a subgroup of the population.
How is it measured?	A wide range of health outcome measures are available for assessing health system performance. Most are highly standardized. The key indicators for measuring health status are: rates of mortality, fertility, and morbidity; life expectancy at birth; self-rated health; and, summary measures such as disability-adjusted life years (DALYs) and quality-adjusted life years (QALYs). Health outcomes can be assessed at the level of the country, state, district or other geographical unit. They can be disease-specific (e.g., prevalence of tuberculosis or hypertension in the population) or life-stage specific (e.g., maternal, infant, and neonatal mortality rate). Broader population-based health status indicators include crude birth and death rates and life expectancy at birth. Measurement of health status need large datasets, as from surveys or records about births and deaths, verbal autopsies, or causes of death.
What data are commonly available?	Most countries have sufficient secondary and administrative data to assess health status without collecting new data specifically for a health system assessment. Health status data can be sourced from vital registration systems, the Demographic and Health Surveys (DHS) (https://dhsprogram.com/) and other population health surveys, global estimations such as the Global Burden of Disease Studies (https://www.healthdata.org/gbd) conducted by the Institute for Health Metrics and Evaluation (IHME), and other national- or state-level administrative data that estimate mortality and morbidity burdens.

	Disease-specific indicators are available in several WHO databases (e.g., the TB database https://www.who.int/teams/global-tuberculosis-programme/ or the HIV database https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/). The OECD does comparisons across countries every other year in its *Health at a Glance* reports (https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2021_ae3016b9-en)
What are the common interpretations?	Population-level measures of health status indicate the population's health as a whole. However, analyses disaggregated by sociodemographic characteristics (focusing on race/ethnicity, age, gender, geographical and rural-urban differences, among others) are necessary to assess the distribution and inequities of health status.
What are the common data gaps?	Most health outcome surveys, including the DHS, focus on infectious diseases and maternal and child health. Data on the non-communicable diseases, mental health, accidents, and injuries that constitute major burdens of disease are usually missing from these surveys, especially in lower-income countries. Vital registration systems, hospital records, and verbal autopsies may not be sufficiently robust or regular to assess mortalities and morbidities.

Final Outcome #2: Financial Risk Protection	
How is it defined?	"Financial risk protection" is achieved when direct payments made to obtain health services do not expose people to financial hardship and do not threaten their living standards. It combines two key questions: (1) What is the extent to which the health system protects people from the financial risks of disease? (2) Do healthcare costs require households to forego other essential goods and services (such as food, housing, or education)?
How is it measured?	Measures of financial risk protection focus on out-of-pocket (OOP) or direct payments made to healthcare providers when goods or services are received. It includes two components. The first is the total amount of money spent in accessing healthcare (which includes the amount of direct health expenditures, e.g., expenses on hospital fees, medicines, diagnostic tests, etc., and the indirect expenditures, e.g., wage loss and travel costs to access health services, informal payments or bribes to access care, etc.). The second component is how the system protects households from the unpredictability (or "shock") of paying for an unplanned health event. One common measure of financial risk protection is "catastrophic health expenditure" (CHE), when OOP spending exceeds a pre-defined share of household income or household consumption spending. The second common measure is "impoverishing health expenditure" (IHE), which measures whether a household's consumption expenditure falls below the poverty line after health spending is

	subtracted. Usually, OOP expenditures exceeding 10% of total household consumption expenditure are considered CHE, and those exceeding 25% are considered IHE.
What data are commonly available?	The data for financial risk protection are generated from surveys in which households report on their spending, total expenditures, and income. If this data is unavailable, analysts may also consider: the OOP costs of a healthcare encounter or episode of illness; distress financing (i.e., whether patients sell assets or borrow funds to cover healthcare costs); and, foregone care due to healthcare costs. Common data sources are: the National Health Accounts (https://apps.who.int/nha/database) undertaken by many countries, national household surveys and administrative data on consumption expenditures and insurance coverage (e.g., the National Sample Surveys in India or the National Survey of Household Consumption and Expenditure in Mexico), and the WHO and World Bank Global Health Expenditure databases (https://databank.worldbank.org/databases/health-financing) that compile data on various health financing and financial risk protection indicators for most countries. DHS data in most countries also includes basic financial indicators including health expenses, household income, and insurance coverage (https://dhsprogram.com/)
What are the common interpretations?	Some financial risk protection metrics make the health system look like it is performing well but fail to consider access to care and the need for health services. For example, CHE and IHE may be low, but it may be because high OOP costs deter people from seeking necessary health services. Analyses of CHE and IHE should also be disaggregated by income level and sociodemographic characteristics in order to generate nuanced assessments and indicate equity.
What are the common data gaps?	In most lower-income countries, a common gap is disaggregated data on OOP expenses (e.g., What households are spending on—is it outpatient or inpatient care? And which components of care: medicines, diagnostics, indirect expenses, etc.?)

Final Outcome #3: Public Satisfaction	
How is it defined?	"Public satisfaction" is the degree to which citizens (or the general public) are satisfied with the services provided by the health system. It pertains to the satisfaction of both users *and non-users* of the available healthcare services. It incorporates peoples' experiences with service provision with broader factors, such as: trust in the health system, confidence that one would receive care if one falls ill, and perceptions about whether the health system needs major changes.
How is it measured?	Public (or citizen) satisfaction is measured primarily by surveys, which ask respondents who are representative of the population of

	interest to report their satisfaction with the health system. These surveys involve face-to-face interviews with individual representatives of households. The response categories almost always use a Likert scale with four or five points. Due to the nature of this outcome, citizen satisfaction is mostly measured for the national or state health system. (Levels below that, such as for individual facilities, are addressed through patient satisfaction assessments, which are discussed under quality of care below).
What data are commonly available?	Some of the most prominent surveys that regularly measure citizen satisfaction are: the Eurobarometer Survey (https://europa.eu/eurobarometer/), which measures public satisfaction among European citizens in 15 member states of the European Union; the Commonwealth Fund's health policy surveys (https://www.commonwealthfund.org/publications/surveys), which measure healthcare consumer satisfaction in selected countries; and, the Gallup World Poll (https://news.gallup.com/poll/4708/healthcare-system.aspx), which measures satisfaction with a range of public institutions (health, education, and justice) across many countries. Several studies across the world have successfully adapted the Commonwealth Fund's International Health Policy Survey questions to measure different aspects of public satisfaction. Additionally, a number of (mostly high-income) countries have questions about satisfaction in national health surveys.
What are the common interpretations?	Public satisfaction is a politically valuable outcome to assess citizens' perceptions of government programs and policies. Approaches to measuring citizen satisfaction are not as standardized as health outcome or financial risk protection measures. Health system analysts must think critically about how satisfaction questions are asked and interpreted in any given context. Because expectations mediate public satisfaction, it can be complicated to assess this outcome. Lower levels of public satisfaction among disadvantaged populations can indicate systemic inequities. However, evidence shows that disadvantaged populations sometimes have lower expectations. Therefore, they may report a higher level of satisfaction with the health system even when objective metrics of service provision, access, quality, financial risk protection, and inequities indicate significant problems. Public satisfaction needs to be interpreted carefully and contextualized to sociocultural realities.
What are the common data gaps?	Public satisfaction data are not measured by national- or state-level surveys in most lower-income countries. In 2002-4, the WHO World Health Survey (https://apps.who.int/healthinfo/systems/surveydata/index.php/catalog/whs) collected data on some public satisfaction-related variables, e.g., the responsiveness of health systems, from multiple countries. However, the findings are now out-of-date. Other studies have assessed public satisfaction and explored concepts such as citizens' trust and

	confidence in the health system across LMICs, but these data have not been collected regularly.

Intermediate Outcome #1: Access	
How is it defined?	Meaningful "access to health care" means that the appropriate healthcare provider or service is supplied in the right place and at the right time to meet the prevailing needs of the population and that the population utilizes the services. Thus, access does not just mean physical availability (of health facilities, healthcare providers, medicines, vaccines, diagnostics or other medical products). It also means that the population knows what is available, seeks out healthcare and uses it to prevent, manage, and treat health conditions.
How is it measured?	Access is measured through both supply-side and demand-side indicators. Data to measure the supply-side (what is offered) aspects of access are collected using facility/provider surveys (which are undertaken at national- or state-level in most countries) or health information systems (HIS) that generate data about the number of facilities and health workers in a given geographical unit. Indicators of the supply side of access include various assessments of physical and human resource *inputs* (including: the ratio of doctors to nurses, the number of hospital beds per 1000 population, or attributes of health care facilities, such as whether they have water, electricity, essential medicines, and equipment). Demand-side aspects of access are measured through large household surveys conducted at national-, state-, or district-level. Indicators for the demand side of access include: utilization of healthcare services, number of fully vaccinated children, number of institutional births, and the percentage of women receiving antenatal care. In some cases, a "care cascade" measure is used—this tracks the entire process of care from the time the individual "enters" the health system, beginning with seeking care, through management and treatment, to rehabilitation. Care cascades are commonly used to measure access for chronic diseases that require continuous use of health services.
What data are needed for assessment?	Access to care is one of the most commonly measured outcomes. Most countries have extensive and regular data on both supply- and demand-side access indicators. Common datasets that are available for most countries are: the Demographic and Health Surveys (DHS) (https://dhsprogram.com/) that capture indicators on utilization of services (e.g., antenatal care, institutional childbirths, vaccinations, etc.); Service Provision Assessments (SPA) that are conducted as part of the DHS (https://dhsprogram.com/methodology/Survey-Types/SPA.cfm) to collect data on supply-side indicators related to physical and human resource attributes; the Service Availability and

	Responsiveness Assessments (SARA) conducted by WHO (https://www.who.int/data/data-collection-tools/service-availability-and-readiness-assessment-(sara)); the World Bank's Service Delivery Indicator Survey (https://datacatalog.worldbank.org/search/dataset/0042030), and health information systems used in most countries, which report number of health facilities, ratios of health personnel, vacancies, etc.
What are the common interpretations?	Access to care measures are useful for assessing the *inputs* invested in the health system. These measures are also closely associated with other outcomes—thus access to care cannot be meaningfully interpreted without also understanding the affordability and quality of the services available. Further, the physical availability of health services does not indicate whether the services are effective or if they are being used by the population. Similarly, uptake/utilization of health services by the population does not indicate that people are receiving high-quality care or that good health outcomes are being produced. As with other indicators, access to care also needs to be disaggregated by sociodemographic characteristics of the population to examine equity.
What are the common data gaps?	While most countries have extensive data on access, these are usually limited to the public sector and formal healthcare providers. Private-sector providers, including informal or unlicensed providers, are rarely included in supply-side facility surveys—even when a majority of the population receives care from private providers. Another major data gap is on care cascades. Most demand-side surveys focus on access at one point in time (or for specific visits), but this does not capture people who need care but either forego or drop out of care, especially in the case of chronic diseases.

Intermediate Outcome #2: Quality of Care	
How is it defined?	"Quality of care" is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes. Quality healthcare has been defined in many ways, but there is growing acknowledgment that it includes three main aspects: 1. *Clinical effectiveness*–providing evidence-based healthcare services to the people who need them while avoiding overuse of inappropriate care and underuse of effective care. 2. *Patient safety*–avoiding causing harm to the people receiving care. 3. *People-centeredness*–providing health care that responds to and respects the preferences, needs, and values of the people who need services.

How is it measured?	Each of the three aspects of quality has its own measurement processes and indicators. 1. *Clinical effectiveness* is measured by comparing the care provided with current evidence on effective diagnostic and treatment guidelines. Assessments of clinical effectiveness measure the extent to which a diagnostic process, diagnosis, or treatment is based on standard guidelines shown to impact health outcomes. Most methods of measuring clinical effectiveness require having clinically-trained data collectors and analysts. Clinical effectiveness can be measured using: clinical vignettes to assess provider knowledge; standardized patients to assess provider practices and "know-do" gaps (that is, the difference between what providers know and what they actually do); direct observations of patient-provider interactions; and reviews/audits of claims data, patient records, charts, and prescriptions that assess the correctness of treatments. Examples of clinical vignettes and standardized patient interview guides can be found here: https://www.ahrq.gov/ncepcr/tools/healthier-pregnancy/presentations/vignettes.html. 2. *Patient safety* is measured by assessing the number of adverse events or errors at a facility. At hospitals, this may include, for example, incidents of patients receiving infusions of the wrong blood type, patient falls during hospital stays, sponges left inside surgical sites, allergic reactions to medicines not recorded in the patient record, etc. The Hospital Survey of Patient Safety Culture (HSOPS) (https://www.ahrq.gov/sops/surveys/hospital/index.html) is a globally validated survey instrument that assesses the perceptions of clinicians and other staff of the culture of safety in their health facilities. It assesses the context and enabling systems that encourage reporting of adverse events. 3. *People-centeredness* is frequently measured using exit interviews and surveys with patients about their experiences with health care. Different surveys are used assess different aspects of the visit, including the patient's satisfaction with the entire visit and/or the provider, perceptions about convenience and physical aspects of the visit (e.g., wait times, privacy, during the consultation, etc.), and their interactions with providers (e.g., time spent with the patient, respectfulness, etc.). Patients typically rate the aspects of the visit on Likert scales. The globally validated patient survey instrument called the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) (https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalHCAHPS) have been adapted and used in several countries.

What data are commonly available?	Robust and regular quality of care data are rarely available for LMICs and LICs; MICs and HICs also have major data gaps. Clinical vignettes and chart review assessments are used by several HICs as part of their health information systems and routine evaluations of physicians. Audits of prescriptions and insurance claims data can be used for assessing quality, although identifying gaps in provider knowledge or know-do gaps is not possible with administrative data. Patient safety data can be generated through internal error reporting systems in hospitals and hospital audits. Patient satisfaction surveys are increasingly becoming common in health systems around the world, including in LMICs.
What are the common interpretations?	Different measurement methods assess different aspects of quality. Clinical vignettes, for example, only assess a provider's knowledge, not their actual practice. As such, vignettes often produce an overestimation of clinical effectiveness. Combining standardized patient interviews with vignettes can be used to indicate both knowledge and practice. Chart reviews, prescription audits, and insurance claims assess provider practice, but without indicating whether the condition was diagnosed correctly. Patient safety data need to be interpreted in the context of the culture and systems for adverse event reporting. A report of zero adverse events does not mean that care is completely safe—it might mean that either providers are not reporting adverse events or that a system for reporting does not exist or is not enforced. People-centeredness is linked to people's expectations of the quality of care. If expectations are low, patients might report high levels of satisfaction even when objective measures indicate poor care quality. E.g., patients from socio-economically disadvantaged groups are more likely to have lower expectations from public services and, therefore, may be more easily satisfied with a healthcare visit. Additionally, patients are not able to judge the clinical quality of care and may use visible markers/proxies as indicators of high quality. E.g., patients might rate a provider who prescribes multiple medicines and diagnostic tests highly even though several of these prescriptions might be unnecessary or even potentially harmful.
What are the common data gaps?	Data on the quality of care is scarce, especially on clinical effectiveness and patient safety. Although there are studies from LMICs on clinical effectiveness, these have been relatively small research studies. Most insurance claims data are not sufficiently disaggregated to assess clinical effectiveness. Patient satisfaction data on aspects like abuse, safety, and corruption are rare in most

	health systems. Additionally, satisfaction ratings are often not weighted or disaggregated by patients' sociodemographic characteristics to assess equity and the effect of lower expectations leading to high ratings.

Intermediate Outcome #3: Efficiency	
How is it defined?	The concept of efficiency is based on the relationship between a health output and the inputs required to produce it. In an efficient system, the amount of input utilized should result in the production of the maximum amount of output that is possible. Efficiency is usually defined in three ways: 1. Technical efficiency (TE) - when the maximum possible output is produced from a given set of inputs. Alternatively, this implies minimizing the amounts of inputs used to produce a given amount of output. TE involves making sure that the right mix of physical inputs, such as personnel, equipment, supplies, and facilities, are used to produce a health output. 2. Price efficiency (PE) - when the maximum possible output is produced at the lowest possible cost of inputs. PE refers to the right mix of monetary inputs used in the production process. PE incorporates the idea of TE since minimization of costs can be achieved by reducing the misuse of inputs as well as reducing unnecessary expenditures. 3. Allocative efficiency (AE) - when more than one output is produced in a health system, the question of how the inputs are distributed among the production of each output becomes relevant. AE captures the extent to which the inputs are being used to produce the correct mix of outputs that maximize health status gains
How is it measured?	Two broad approaches have been commonly used to measure the efficiency of a health-providing unit – ratio-based and frontier analysis. 1. Ratio-based measures are the most common efficiency measure. It is a ratio of a health system input to the output that it produces. A ratio indicates the resource used per unit of health system output; the greater the ratio, the more efficient the health-providing unit. TE is calculated by dividing any measure of health output by the physical unit of input. E.g., a physician's productivity is calculated as the total number of hours the physician spent in patient care (input) divided by the number of visits (output), which is a measure of TE. Similarly, the generic prescribing rate, the total number of generic

	medications prescribed in a day divided by the total number of patients examined in a day, is also a TE measure. Using input costs as the denominator generates an indicator of PE. Cost per episode is an example where the output is a bundle of services, including visits, medications, procedures, and urgent care services provided for the care of a specific illness, and the input is the monetized total costs of the care. 2. Frontier analysis measures are based on the production function, the relationship between the health inputs and outputs, and the efficiency of a health-providing unit involves comparing its actual performance with the optimal performance located on the production frontier. Since the true frontier is unknown, an empirical approximation is needed. That is, efficiency scores for each unit are based on the frontier function estimation, followed by the distance of the unit from the efficient frontier. These methods can be used to obtain technical, price, or allocative efficiency scores depending on the function – production, cost, profit – that is estimated. There are different ways to estimate the production function, but two methods have been typically used in the study of health systems – data envelopment analysis and stochastic frontier analysis.
What data are commonly available?	Data on health inputs and outputs are needed to calculate any measure of efficiency, and the information is usually available at different levels of the health system (national, state, district, facilities, providers, and households). The choice of the data source depends on the unit of analysis, e.g., is efficiency calculated at the country level or facility level. Common data sources include health information systems that provide a count of the number of health inputs like the number of physicians/nurses, hospital beds, medical equipment, etc. Budget and expenditure data can be used to calculate the cost inputs; other data sources like the Service Provision Assessment (SPA) (https://dhsprogram.com/methodology/Survey-Types/SPA.cfm) could be used to assess the health system inputs and service provision outputs (e.g., total number of physician hours spent in patient care). If the efficiency analysis is focused on health outcomes, then health surveys like the ones mentioned under Health Status, like the Demographic and Health Surveys (DHS) (https://dhsprogram.com/), or insurance claims data can be used.
What are the common interpretations?	Ratio-based efficiency indicators are useful when the intent of the focus is on a particular input or part of the production process. A ratio also allows for comparison across health systems or health-providing units of different sizes/levels. However, each ratio provides a very narrow view of efficiency without accounting for the many interdependent inputs that go into the production of multiple health outcomes. A key advantage of the frontier-based efficiency measures is that they account for multiple inputs and also

	allow for the statistical testing of hypotheses on the relationship between external factors (unit ownership, competition, etc.) and the estimated efficiency scores. However, these measures are complicated to implement and require specialized software and econometric training among analysts.
What are the common data gaps?	Health facility-level data on inputs and potential outputs are often difficult to obtain. Facilities, especially for-profit providers, may not be willing to share this information. Even when available, most input data focuses on hospitals, and there is very scarce data about primary care providers or individual providers. The other gap is in meaningfully linking the inputs to outputs. While health outcome data are available, they are determined by various factors, including those outside the health system. This makes it difficult to attribute outputs to the inputs in efficiency assessments. Usually, health service provision indicators are used as a proxy, but disaggregated data on these indicators are difficult to obtain in many lower-income country health systems.

Appendix 3-2: Additional resources on assessing health system outcomes

Financial Risk Protection
Video training session: https://www.youtube.com/watch?v=2AHR7GN3Omw
Reading:

- Wagstaff A, Flores G, Hsu J, Smitz M-F, Chepynoga K, Buisman LR, van Wilgenburg K, Eozenou P. 2018. Progress on catastrophic health spending in 133 countries: A retrospective observational study. The Lancet Global Health, 6(2), e169–e179. https://doi.org/10.1016/S2214-109X(17)30429-1

Public Satisfaction
Video training session: https://www.youtube.com/watch?v=UxE1CU23Mgk
Readings:

- Blendon RJ, Benson J, Donelan K, Leitman R, Taylor H, Koeck C, Gitterman D. 1995. Who Has The Best Health Care System? A Second Look. Health Affairs 14(4), 220–230. https://doi.org/10.1377/hlthaff.14.4.220
- Blendon RJ, Kim M, Benson JM. 2001. The Public Versus The World Health Organization On Health System Performance. Health Affairs 20(3), 10–20. https://doi.org/10.1377/hlthaff.20.3.10

Access
Video training session: https://www.youtube.com/watch?v=bR4mAq3o4J0

Quality of care
Clinical effectiveness
Video training session: https://www.youtube.com/watch?v=CFaTmPjG5KQ
Readings:

- Das J, Holla A, Das V, Mohanan M, Tabak D, Chan B. 2012. In Urban And Rural India, A Standardized Patient Study Showed Low Levels Of Provider Training And Huge Quality Gaps. Health Affairs, 31(12), 2774–2784. https://doi.org/10.1377/hlthaff.2011.1356
- Holla A. 2013. Measuring the Quality of Health Care in Clinics. World Bank Group. https://www.globalhealthlearning.org/sites/default/files/page-files/Measuring%20Quality%20of%20Health%20Care_020313.pdf (Accessed 2 August 2023)
- Kruk M E, Gage AD, Arsenault C, et al. 2018. High-quality health systems in the Sustainable Development Goals era: Time for a revolution. Lancet Global Health, 6(11), e1196–e1252. https://doi.org/10.1016/S2214-109X(18)30386-3

Patient Safety
Video training session: https://www.youtube.com/watch?v=3c4KXF4h6ik
Readings:

- DiCuccio MH. 2015. The Relationship Between Patient Safety Culture and Patient Outcomes: A Systematic Review. Journal of Patient Safety, 11(3):135–142. https://doi.org/10.1097/PTS.0000000000000058
- Fontana G, Flott K. Dhingra-Kumar N, Durkin M, Darzi A. 2019. Five reasons for optimism on World Patient Safety Day. The Lancet, 394(10203):993–995. https://doi.org/10.1016/S0140-6736(19)32134-8
- Jha AK, Larizgoitia I, Audera-Lopez C, Prasopa-Plaizier N, Waters H, Bates DW. 2013. The global burden of unsafe medical care: Analytic modelling of observational studies. BMJ Quality & Safety, 22(10), 809–815. https://doi.org/10.1136/bmjqs-2012-001748

People centeredness

Video training session: https://www.youtube.com/watch?v=mkXkZ6Xwpo8

Reading:

- Larson E, Sharma J, Bohren MA, Tunçalp Ö. 2019. When the patient is the expert: Measuring patient experience and satisfaction with care. Bulletin of the World Health Organization, 97(8), 563–569. https://doi.org/10.2471/BLT.18.225201

Efficiency

Video training session: https://www.youtube.com/watch?v=oRDLX2QkHHs

Readings:

- Yip W, Hafez R. 2016. Improving health system efficiency: reforms for improving the efficiency of health systems: lessons from 10 country cases. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/185989
- Hafez R, ed. 2020. Measuring Health System Efficiency in Low- and Middle-Income Countries: A Resource Guide. Joint Learning Network for Universal Health Coverage. https://www.jointlearningnetwork.org/resources/resource-guide-for-measuring-health-system-efficiency-in-low-and-middle-inc/
- McGlynn EA. 2008. Identifying, Categorizing, and Evaluating Health Care Efficiency Measures. Final Report (Publication No. 08-0030). Rockville, MD: Agency for Healthcare Research and Quality. https://library.ahima.org/PdfView?oid=81708 (Accessed 2 August 2023)