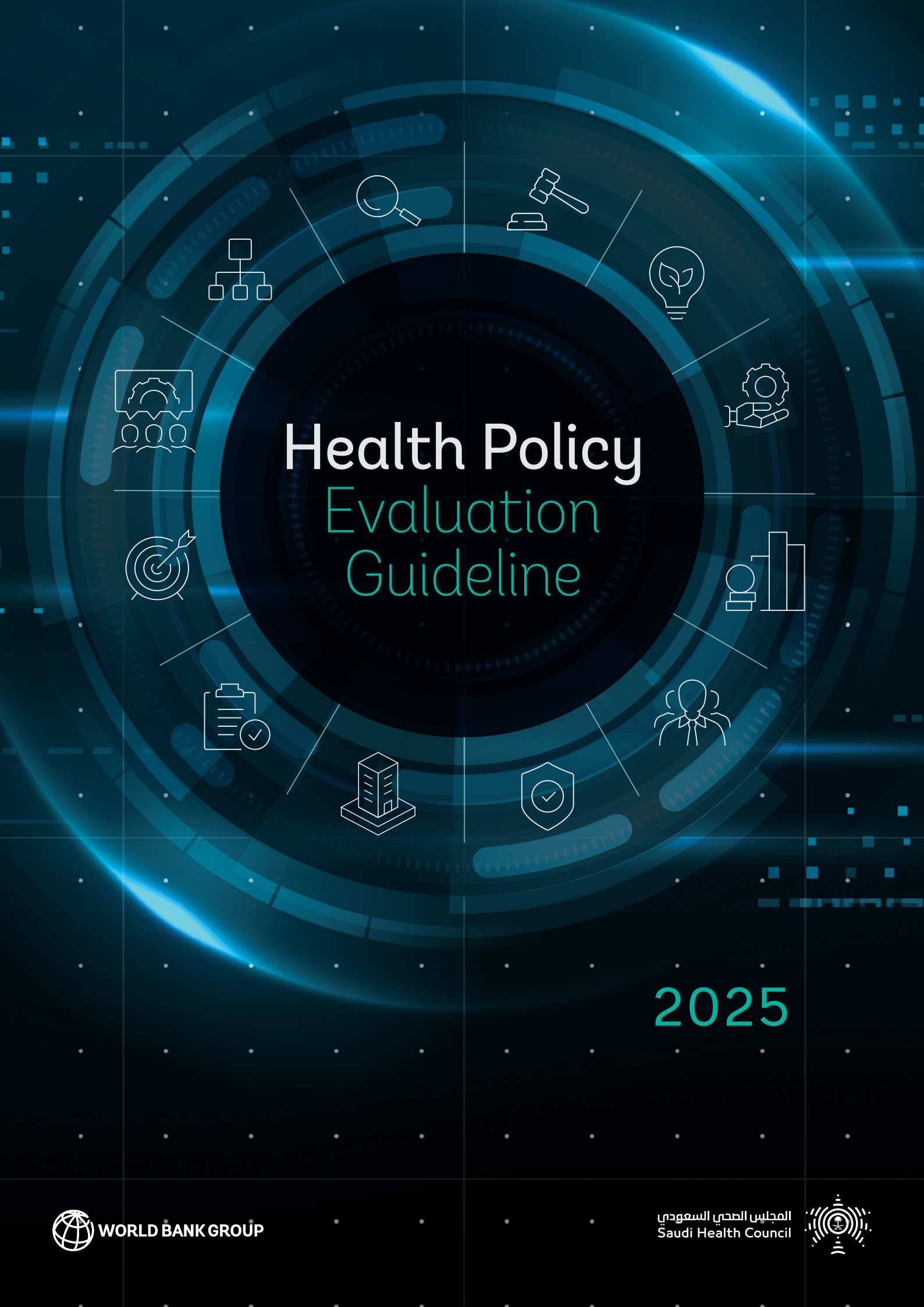

Health Policy
Evaluation
Guideline
2025
WORLD BANK GROUP
المجلس الصحي السعودي
Saudi Health Council

Health Policy Evaluation Guideline

Saudi Health Council
Kingdom of Saudi Arabia

2025

Acknowledgment

This guideline was jointly developed by the Saudi Health Council (SHC) and the World Bank (WB) under the Reimbursable Advisory Services (RAS) program between the Kingdom of Saudi Arabia (KSA) and the WB. It was prepared under the strategic direction of Khalid Almoteiry (SHC) and Volkan Cetinkaya (WB). Core contributions were provided by the SHC team—Shahad Al-Homidi, Ahlam Alshehri, Ali Alqaysi, Rana Saber, Khalid Farhan, Abdulaziz Alrabiah, and Mushabab Al-Asiri; and the WB team—Parviz Ahmadov and Isa Aliyev. Richard Crabbe provided editorial services and Dania Kibbi handled design and production.

We are grateful to the National Advisory Team for their technical guidance and oversight throughout the process. The team included Bader Alamer (Health in All Policies General Secretariat); Ahmed Almokharshum (Ministry of Health); Faisal Sunaid (Saudi Food and Drug Authority); Muhannad Almughthim (Saudi Commission for Health Specialties); Dalia Alghamdi (Ministry of Defense); Thamir Alsaiary (Ministry of National Guard – Health Affairs); Haitham Bader (Ministry of Interior); Ziyad Rayes (King Faisal Specialist Hospital and Research Centre); Ziyad Almalki (Ministry of Education); and Razan Aljaser (Council of Health Insurance).

We also thank H.E. Dr. Nahar Al-Azemi (Secretary General, SHC); and Michele Gragnolati (Practice Manager for Health, Nutrition, and Population, Middle East and North Africa Region, WB) for their overall support and leadership.

This guideline is a technical product of the SHC–World Bank RAS program. It is intended to support institutional capacity-building and the standardization of health policy evaluation practices in KSA. As a living document, it may be further refined over time through implementation experience and stakeholder feedback.

Table of Contents

List of Tables

List of Figures

Glossary of key terms

This section presents key terms and definitions used throughout this Health Policy Evaluation Guideline. It aims to establish a shared understanding among policy makers, evaluators, researchers, and practitioners involved in designing, implementing, or using health policy evaluations in the Kingdom of Saudi Arabia.

Terms	Definition
Adaptive Evaluation	A flexible evaluation approach that evolves in response to real-time implementation challenges and emerging findings.
Assumptions	The conditions or factors believed to hold true in the Theory of Change or logic model, which if incorrect, can affect the success or validity of the evaluation.
Attribution vs. Contribution	Attribution assigns direct causal responsibility to a policy for observed changes. Contribution recognizes the policy as one of several influencing factors in a complex system, without exclusive causal claims.
Baseline Data	Data collected before policy implementation, serving as a reference point to assess changes and policy impact.
Benchmarking	Comparing policy performance against standards, past performance, or peer jurisdictions to assess effectiveness and identify improvement opportunities.
Causal Attribution Methods	Techniques used to determine whether observed effects can be causally linked to a policy, including randomized control trials, difference-in-differences, and instrumental variables.
Causal Chain	A logical sequence showing how policy inputs and activities lead to outputs, outcomes, and impacts. Often visualized through logic models.
Cost-Effectiveness Analysis (CEA)	A form of economic evaluation that compares the relative costs and outcomes of different interventions to assess value for money.
Counterfactual	An estimate of what would have occurred in the absence of the policy, used in causal attribution and impact evaluation designs.
Data Governance	The framework of rules, procedures, and ethical principles that guide the collection, management, protection, and use of data in health evaluations.
Disability-Adjusted Life Year (DALY)	A composite measure of disease burden that combines years of life lost due to premature death (YLL) and years lived with disability (YLD). One DALY represents one lost year of healthy life.
Disaggregation	The breakdown of evaluation data by subgroups – (for example, age, gender, region) to assess differential effects and promote equity.
Equity Assessment	The analysis of how a policy affects different population groups, particularly vulnerable or underserved populations, to promote fairness.
Ethical Clearance	Approval from an ethics review body that ensures that evaluations meet national and international human subject protection standards.
Evaluation	A systematic, objective, and time-bound process of assessing a health policy's design, implementation, and results. It provides evidence on whether the policy was appropriately formulated, implemented as intended, and effective in achieving its intended outcomes or impacts.
Evaluation Methods	The tools and techniques used to gather evidence in evaluations, including qualitative methods such as interviews; quantitative methods, for example, surveys; or mixed methods.
Evaluation Type	The specific focus of an evaluation, including regulatory impact analysis, process, outcome, impact, or economic evaluation, aligned with different stages of the policy lifecycle.
Evaluation Utilization	The process of applying evaluation findings to improve policy design, implementation, scaling, or de-implementation decisions.
External Factors	Influences outside the policy – for example, economic shocks, pandemics, or parallel reforms that may affect its outcomes or impacts – which must be considered in evaluation design.

Field Testing (of Indicators)	Piloting indicators in actual settings to evaluate their clarity, feasibility, and reliability before full implementation.
Health Policy	Decisions, plans, or actions undertaken by governments or institutions to achieve public health objectives. These may be preventive, curative, or systemic and operate at various levels of the health system.
Health Policy Cycle	The cyclical process of initiating, designing, implementing, evaluating, and revising policies to adapt to emerging health system needs.
Health Policy Evaluation	A systematic and objective assessment of a health policy's design, implementation, and outcomes, aimed at determining its relevance, effectiveness, efficiency, impact, and sustainability.
Health Technology Assessment (HTA)	A comprehensive evaluation of the clinical, economic, and social aspects of health technologies to guide policy decisions.
Impact	The long-term, systemic changes a policy aims to produce, such as reduced mortality, improved health equity, or increased life expectancy. Impact is influenced by multiple external factors.
Implementation Barriers	Contextual, organizational, or systemic obstacles that prevent a policy from being delivered as intended – for example, lack of staff, legal constraints, or technology failures.
Indicators (Process, Output, Outcome, Impact)	Measurable variables used to monitor and evaluate different aspects of a policy's performance across the results chain.
Informed Consent	A formal process ensuring that evaluation participants understand and voluntarily agree to participate in line with ethical standards.
Inputs	The financial, human, and material resources mobilized to support policy implementation – for example, funding, staff, and equipment. Inputs are the starting point of the logic model.
Logic Model	A visual or narrative tool that maps the relationships between inputs, activities, outputs, outcomes, and impacts, guiding evaluation design.
Logical Framework Approach (LFA)	A structured planning and evaluation methodology using a hierarchy of objectives, indicators, assumptions, and means of verification.
Mixed-Methods Evaluation	An approach that integrates both qualitative and quantitative data collection and analysis to provide a more complete understanding of policy performance and impact.
Monitoring	A continuous and routine process of tracking a policy's inputs, activities, and outputs using predefined indicators. It supports operational oversight, identifies implementation issues, and enables timely adjustments.
Outcomes	The short- to medium-term effects of the policy, such as increased coverage, improved behaviors, or better quality of care. These result from outputs and precede long-term impact.
Outputs	The immediate results of policy activities, such as services delivered, training conducted, or materials produced. Outputs are tangible and within the control of the implementing agency.
Performance Standards	Established benchmarks or thresholds against which actual policy results are assessed.
Policy Impact Assessment	A type of evaluation that measures long-term, system-level, and population-wide effects of a policy, including unintended consequences.
Policy Implementation Fidelity	The extent to which a policy is delivered as intended, critical for understanding deviations that may affect results.
Power Calculation	A statistical process used to determine the appropriate sample size needed to detect a policy effect with a given level of confidence.
Process vs. Structural Indicators	Process indicators measure actions taken, such as services delivered; structural indicators assess system capacity, for example, staff and infrastructure.
Quality-Adjusted Life Year (QALY)	A metric that combines life expectancy and quality of life to assess the value of health interventions in economic evaluations.

Quasi-Experimental Design	An impact evaluation approach that compares intervention and non-intervention groups without randomization, used when Randomized Controlled Trials (RCTs) are not feasible.
Real-World Evidence (RWE)	Evidence derived from routine clinical or administrative data (rather than trials), increasingly used for evaluation in applied settings.
Realist Synthesis	An evaluation approach that explores how and why policies work in specific contexts, focusing on context–mechanism–outcome (CMO) configurations.
Reliability	The degree to which evaluation methods yield consistent and replicable results across repeated applications or evaluations.
Sensitivity Analysis	A technique used to test how the results of an evaluation change when key assumptions, inputs, or parameters are varied. Common in economic evaluation.
Stakeholder Engagement	The process of identifying and involving individuals or groups affected by, or influential to, the policy and its evaluation to ensure relevance, trust, and use of findings.
Target Groups	The individuals, populations, or health system actors for whom the health policy is intended—whether as direct beneficiaries (e.g., patients, communities) or as implementing agents (e.g., healthcare providers, facilities, or administrative bodies).
Theory of Change (ToC)	A structured explanation of how and why a policy is expected to achieve its goals, outlining assumptions and causal linkages.
Triangulation	A method of validation that uses multiple sources, methods, or perspectives to increase the credibility and reliability of evaluation findings.
Type I and Type II Errors	Statistical errors in hypothesis testing: Type I error occurs when an evaluation incorrectly concludes that a policy had an effect (false positive). Type II error occurs when it fails to detect a real policy effect (false negative).
Use Case (in evaluation)	A defined policy scenario or decision-making context in which a specific evaluation method, tool, or framework is applied to address an operational or strategic need.
Validity (Internal and External)	This refers to the extent to which the observed outcomes can be attributed to the policy, free from confounding factors. External validity refers to the degree to which results can be generalized to other settings, populations, or contexts.

Foreword

In line with the Kingdom of Saudi Arabia's Vision 2030 and the national commitment to building a high-performing, transparent, and people-centered health system, the Saudi Health Council is proud to present this Health Policy Evaluation Guideline. As the health sector undergoes transformative reforms, the ability to assess the design, implementation, and outcomes of health policies with rigor and consistency becomes more critical than ever.

This guideline serves as a foundational tool to support evidence-informed policymaking. It offers a practical and structured approach for evaluating health policies across their lifecycle—from early planning and implementation to long-term impact assessment. By embedding evaluation within the health policy process, we can ensure that decisions are data-driven, resources are used efficiently, and programs are continuously refined to deliver the best possible outcomes for the Saudi population.

The guideline draws on global best practices while remaining firmly grounded in the Saudi context. It provides clarity on the types of evaluations that can be conducted, the steps to follow, and the standards that must be upheld to ensure transparency, accuracy, and accountability. It also highlights the importance of inclusive stakeholder engagement and the use of logic models to clarify assumptions and causal pathways.

We encourage all policy makers, public health professionals, healthcare institutions, and partners to use this guideline as a standard reference when designing, implementing, or evaluating health policies. By doing so, we move one step closer to realizing a health system that is not only efficient and sustainable but also equitable and responsive to the evolving needs of our people.

On behalf of the Saudi Health Council, I commend all those who contributed to this important work and look forward to its widespread adoption across the health sector.

Dr. Nahar Al-Azemi, MD
Secretary General
Saudi Health Council

01

Executive Summary

The Health Policy Evaluation Guideline for the Kingdom of Saudi Arabia (KSA) is a national reference developed by the Saudi Health Council (SHC) to institutionalize rigorous, context-specific, and systematic evaluation practices across the health sector. Developed in alignment with Saudi Arabia's Vision 2030 and its commitment to transparent, evidence-informed governance, the guideline is designed to support policy makers, regulators, evaluators, and implementing partners in optimizing the design, implementation, and impact of health policies. It offers a unified approach that integrates the best international practices with local realities, aiming to enhance accountability, resource efficiency, and the effectiveness of policy interventions.

This guideline defines health policy evaluation as the systematic and objective assessment of a policy's design, implementation, and results. It recognizes evaluation not as a stand-alone technical activity, but as a strategic function that generates evidence for learning, strengthens public accountability, informs iterative improvements, and enhances the legitimacy and performance of health policy decisions. By embedding evaluation throughout the policy life cycle, the guideline enables stakeholders to assess what works, under what conditions, and for whom.

The guideline identifies five core criteria for evaluating health policies:

1. **Relevance**	Assesses alignment with current health needs and priorities, ensuring that the policy addresses pressing challenges within KSA's health, social, economic, and cultural context.
2. **Effectiveness**	Measures the achievement of intended goals, analyzing outcomes and identifying success factors or barriers.
3. **Efficiency**	Evaluates use of resources – for example, financial, human, time – against outcomes, ensuring cost-effectiveness and value for money.
4. **Sustainability**	Examines the policy's capacity to maintain benefits over time, adapting to changing conditions and resource availability.
5. **Impact**	Assesses broader effects—intended and unintended—on population health, equity, and health system performance, as well as associated social and economic outcomes, including changes in social determinants of health and economic burden or productivity.

To operationalize evaluation across the health policy life cycle, the guideline presents five evaluation approaches:

1. **Formative evaluation**	Conducted during policy design or pre-implementation, often before large-scale rollout, to assess feasibility, acceptability, and appropriateness.
2. **Process evaluation**	Examines how a policy is implemented and whether it is being delivered as intended.
3. **Outcome evaluation**	Measures whether the policy has achieved its short- to medium-term objectives and delivered intended services to the target population.
4. **Impact evaluation**	Assesses whether the policy produced long-term, systemic changes in health status or equity, using rigorous causal attribution methods.
5. **Economic evaluation**	Evaluates cost-effectiveness, cost-benefit, or cost-utility to ensure that resources are used efficiently and equitably.

The guideline introduces a six-step process for conducting health policy evaluations, adapted from international frameworks and best practices, and tailored to the context in KSA:

1. **Identify and engage stakeholders**	Mapping and involving actors who are affected by, implement, or will use the evaluation results, ensuring relevance and ownership.
2. **Describe the health policy**	Articulating the policy's goals, target groups,[1] resources, activities, and intended outcomes using logic models.
3. **Design the evaluation**	Selecting appropriate evaluation types, formulating questions, and choosing methodological approaches—quantitative, qualitative, or mixed.
4. **Collect data**	Gathering credible and context-relevant data through routine systems, surveys, or administrative records.
5. **Analyze and justify conclusions**	Synthesizing findings through triangulation to draw reliable, evidence-based conclusions.
6. **Use and disseminate findings**	Ensuring that evaluation results inform policy adjustments, public accountability, and institutional learning.

Throughout this process, the guideline emphasizes adherence to four evaluation standards:

1. **Utility**	Serving the information needs of decision-makers and stakeholders.
2. **Feasibility**	Being realistic given time, budget, and data availability.
3. **Propriety**	Upholding ethical standards and stakeholder rights.
4. **Accuracy**	Ensuring findings are technically sound, valid, and trustworthy.

To ensure that the guideline is grounded in real-world experience, **the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners** was used as a sample. The data and findings presented in this guideline are illustrative and based on assumptions made by the SHC team, rather than actual data. This policy, which expanded malpractice insurance coverage to 18 non-physician specialties, served not merely as a case study but as a foundational reference throughout the guideline. The evaluation findings were used to structure and validate the six-step process itself—informing the development of stakeholder engagement typologies, logic model construction, indicator design, data collection procedures, and approaches for analyzing and using findings.

By codifying a unified, evidence-based approach to health policy evaluation and demonstrating its application through a nationally relevant policy, this guideline equips KSA's health institutions to assess what works, under what conditions, and for whom. It establishes evaluation as a routine and strategic component of good governance—contributing to more effective health policies, better outcomes for the population, and a resilient, more accountable health system.

1 In this guideline, "target groups" refer to the individuals, populations, or health system actors for whom the health policy is intended, whether as direct beneficiaries – for example, patients and communities – or as implementing agents, such as healthcare providers, facilities, or administrative bodies.

02

Introduction

Health policy evaluation is essential for ensuring that the Kingdom of Saudi Arabia's (KSA) health policies are well-designed, effectively implemented, and aligned with national priorities, including Vision 2030 and international commitments. Vision 2030 outlines an ambitious national transformation agenda, under which the Kingdom is undertaking major healthcare reforms that emphasize a data-driven approach—leveraging performance indicators and measurable outcomes to support continuous monitoring, evaluation, and accountability.[2,3] This strong foundation enables evidence-based decision-making and fosters sustained policy improvement.

This guideline aims to provide a structured framework for the systematic assessment of health policies and to equip policy makers, evaluators, and stakeholders with practical tools to enhance policy impact and societal well-being. It is designed to support KSA in addressing its unique health system challenges while promoting efficient resource use and accountability in the delivery of intended health outcomes.

This guideline applies to a broad spectrum of health policies, including public health initiatives, healthcare regulations, and interventions aimed at improving population health. It offers practical strategies to evaluate health policies across their life cycle—from initial design and implementation to mid- or long-term outcomes—is organized into the following sections:

1. What is a health policy evaluation?	Defines the concept and underscores its significance in the KSA context.
2. Why evaluate health policies?	Explores the dual purposes of learning and accountability that drive the evaluation process.
3. When to evaluate health policies?	Provides guidance on optimal timing across the policy lifecycle.
4. What are the main types of health policy evaluation?	Introduces evaluation types and their applications.
5. How to conduct health policy evaluations in practice?	Details step-by-step methodologies tailored to KSA.

Designed for policy makers, public health professionals, evaluators, and academics in KSA, this guideline delivers actionable insights and methodologies. It contributes to the development of robust, evidence-informed health policies that advance KSA's national health objectives.

2 KSA Vision 2030, *Health Sector Transformation Program*. https://www.vision2030.gov.sa/en/explore/programs/health-sector-transformation-program.
3 Ministry of Health. *Healthcare Transformation Strategy*. Retrieved from https://www.moh.gov.sa/en/Ministry/vro/Documents/Healthcare-Transformation-Strategy.pdf.

03

What is health policy evaluation?

In this guideline, *"health policy" broadly encompasses organized actions, initiatives, or interventions—whether preventive, curative, or systemic—aimed at improving public health or healthcare at local or national levels in KSA*. To ensure clarity and consistency, this guideline establishes a standardized vocabulary for evaluators, healthcare professionals, decision-makers, academicians, and the public, fostering transparent communication throughout the evaluation process.

Based on the international best practices, this guideline defines the health policy evaluation as *"a structured and objective assessment of an ongoing or completed health policy, focusing on its design, implementation, and results. The aim is to evaluate its relevance, efficiency, effectiveness, impact, and sustainability, as well as its overall significance."*[4]

As defined above, this guideline emphasizes evaluating five key aspects of a health policy:[5]

1. **Relevance:** Assesses alignment with current health needs and priorities, ensuring the policy addresses pressing challenges within KSA's health, social, economic, and cultural context.
2. **Effectiveness:** Measures the achievement of intended goals, analyzing outcomes and identifying success factors or barriers.
3. **Efficiency:** Evaluates resource use (e.g., financial, human, time) against outcomes, ensuring cost-effectiveness and value for money.
4. **Sustainability:** Examines the policy's capacity to maintain benefits over time, adapting to changing conditions and resource availability.
5. **Impact:** Assesses broader effects—intended and unintended—on population health, equity, and health system performance, as well as associated social and economic outcomes, including changes in social determinants of health and economic burden or productivity.

To ensure clarity for evaluators and other stakeholders, this guideline also differentiates the concepts of *monitoring* and *evaluation*, which are two complementary yet distinct activities:

1. **Monitoring:** An ongoing process that tracks policy performance using predefined indicators. It provides real-time data for project managers and stakeholders to identify challenges, adjust health policy, and ensure accountability in resource use.
2. **Evaluation:** An episodic, systematic assessment of a policy's design, implementation, and outcomes. It analyzes relevance, efficiency, effectiveness, impact, and sustainability to inform strategic planning and policy refinement.

To guide evaluators in applying monitoring and evaluation, Table 1 summarizes the differences.

Table 1 Key Differences Between Policy Monitoring and Evaluation

Criteria	Policy Monitoring	Policy Evaluation
Nature	Ongoing, operation-focused oversight	Episodic, strategy-focused assessment
System Suitability	Tracks broad, anticipated issues	Addresses specific policy questions
Data Collection	Routine, predefined measures	Customized methods for evaluation goals
Attribution	Links actions to effects via indicators direct links to effects	Analyzes causal links systematically
Resource Allocation	Uses existing infrastructure	Requires dedicated resources
Information Use	Real-time data for management	Findings for future policy design

Source: Adapted from OECD 2020, and World Bank 2009.[6]

4 OECD. 2020. *Improving Governance with Policy Evaluation: Lessons From Country Experiences*. OECD Public Governance Reviews. Paris: OECD Publishing. https://doi.org/10.1787/89b1577d-en.
5 OECD. 2020. *Improving Governance with Policy Evaluation: Lessons From Country Experiences*.
6 Görgens, M., and Jody Zall Kusek. 2009. Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit. Washington, DC: World Bank. https://hdl.handle.net/10986/2702.

04

Why evaluate health policies?

Health policy evaluation serves two core purposes – learning and accountability[7] – which are essential to ensure policy's effectiveness and efficiency, as well as its alignment with KSA strategic health goals incorporated in the "Vision 2030" document.

Under the "learning" function, an evaluation generates evidence to refine policies, drives continuous improvement, and guides future initiatives. The learning function contributes to the following key aspects in health policy design and implementation:

- **Managing risks** assesses whether health policies perform as intended, and reduces implementation uncertainties.
- **Enhancing performance** identifies the opportunities to optimize health policies through using early findings to boost outcomes.
- **Guiding decisions** informs whether to sustain, adjust, or end policies, and improves resource allocation.
- **Shaping future policies** reveals what works, for whom, and why, preventing past errors and strengthening program design.

Under the "accountability" function, evaluations ensure that people implement policies more responsibly, deliver value, maintain trust, and give legitimacy to policies. The accountability function specifically contributes to the following key aspects in a health policy design and implementation:

- **Transparency in spending** verifies if the public funds are used efficiently, and benefits reach citizens.
- **Regulatory compliance:** an evaluation confirms policy's adherence to the national laws and regulations, thereby reinforcing public confidence.
- **Public reporting** openly shares the findings, which demonstrates the success or failure of a policy.

In practice, evaluations are driven by the following main reasons:

- **Formal requirements:** Legal mandates, as in the case of the Medical Malpractice Insurance Policy in KSA, or regulatory roles as in the case of the Saudi Health Council's mandate to review policies.
- **Cabinet oversight:** Guidelines and decisions mandate assessing the financial, economic, and social impacts of policies.
- **Government priorities:** Policies identified as government priorities in national plans or programs – for example, Health Sector Transformation Program (HSTP) initiatives.
- **International commitments:** International agreements or commitments such as the Sustainable Development Goals.
- **Selective evaluations:** Only handle certain policies, due to time and cost constraints – for example, the Medical Malpractice Insurance Policy. The following key criteria could guide decision-makers or evaluators to prioritize the policies to be evaluated:
    - » **Relevance to strategic goals:** Policies that contribute directly to overarching frameworks—such as the KSA Vision 2030 or the Health Sector Transformation Program—should be prioritized because evaluating them can provide critical insights into progress toward these high-level goals. Evaluating such policies ensures that resources are directed toward what matters most at the strategic level.
    - » **Magnitude of impact:** The extent to which a policy affects the population is a key consideration. Policies with wide-reaching effects, or those targeting high-need or vulnerable groups, are more likely to generate meaningful outcomes. Evaluating such policies provides information on how they are improving lives and whether they are achieving their intended impact.
    - » **Level of investment:** High-cost or resource-intensive policies warrant evaluation to ensure that investments are yielding expected results. This includes financial investments, human resources, infrastructure, or administrative commitment. Evaluation helps to assess cost-effectiveness and informs decisions on continuing or scaling back such policies.

7 HM Treasury. 2020. *Magenta Book: Guidance for Evaluation*, pages. 9–10. https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT_Magenta_Book.pdf.

- **Stage of implementation:** Evaluations are most valuable when the policy is at a stage where results can influence implementation—such as during a pilot phase, at mid-term, or when preparing for scale-up. Evaluating too early may not yield results, while evaluating too late may limit the potential for improvement.
- **Feasibility of evaluation:** Even important policies may be deprioritized if evaluation is not feasible. Factors such as the availability of data, access to stakeholders, and appropriateness of methods must be considered. Evaluation should only proceed when it can be conducted rigorously and within the available timeframe and resources.
- **Risks and controversy:** Policies that are politically sensitive, under scrutiny, or potentially associated with adverse outcomes should be prioritized for evaluation. These evaluations can enhance transparency, build public trust, and provide clarity around the policy's effects.
- **Timing and windows of opportunity:** Evaluation should be aligned with key decision-making moments, such as budget planning, policy renewal, or strategic reviews. Prioritizing evaluations that can inform upcoming decisions ensures that findings are timely and actionable.

05

When to evaluate health policies?

Evaluation occurs across the following stages, each targeting a distinct component of the health policy lifecycle:

1. **Before** a policy is fully formed, it is important to use evaluation to help shape its design and how it will be implemented. Using existing evaluation evidence or working through the Theory of Change of the health policy, typically developed during the policy design stage.[8]
   - 1.1. **Input level:** Assesses whether a policy is feasible, appropriate, and acceptable before it is fully implemented. It can include process and outcome measures.

2. **During** implementation provides the greatest opportunity for the evaluation to influence decisions and to help ensure a policy can realize its intended benefits. During implementation, evaluations will typically look at gaining evidence about the efficacy of the policy's design, its implementation, and emerging outcomes.
   - 2.1. **Process level:** Assess the actions and activities undertaken during the implementation of a policy. This involves monitoring the efficiency of healthcare delivery processes and the type or timeliness of a policy.
   - 2.2 **Output level:** Measure the initial results directly associated with the healthcare services delivered by a policy. This includes evaluating patient volumes, the healthcare benefits provided, public health actions, or other service delivery outputs that the policy aims to produce.

3. **After** a policy has been implemented, the entire policy can be examined looking at outcomes and impacts.
   - 3.1 **Outcome level:** Analyze a policy's medium-term effects on health outcomes. This could include assessing improvements in access to health services, reductions in disease incidence, changes in health behaviors, or enhancements in public health knowledge attributable to the policy.
   - 3.2 **Impact level:** Examine the long-term impact of policy on population health. This includes evaluating the overall effectiveness of the policy in achieving public health goals—such as reducing mortality rates, improving quality of life, or promoting health equity across population groups—as well as its broader social and economic effects, including changes in social determinants of health, economic burden, and productivity.

8 For more details on the Theory of Change in policy development, see *The Health Policy Maker's Manual: Integrating Data and Evidence*. 2024. https://shc.gov.sa/Arabic/Documents/The%20Health%20Policy%20Makers%20Manual%20-%20KSA%20-2024.pdf.

06

What are the types of health policy evaluation?

A type of health policy evaluation refers to a specific approach used to assess various aspects of a health policy. The types of evaluation are tailored to address targeted questions[9] about the policy's design, implementation, outcomes, or overall impact. The selection of an evaluation type depends on the level of the policy evaluated and the specific objectives of the evaluation.

Each type of evaluation provides unique insights: enables policy makers and evaluators to determine how effectively a policy is functioning; assesses its effects; and identifies areas for improvement. These insights are crucial for informed decision-making and ensuring that health policies achieve their intended goals.

In some cases, particularly during policy formulation, **Regulatory Impact Assessment (RIA)** may be used as an ex-ante evaluation tool to assess the potential health, social, and economic impacts of proposed policies. RIA supports evidence-informed policy making by systematically analyzing policy options and their likely consequences—see Annex 1 for more on RIA.

The common types of health policy evaluation, along with *what they show, when to use them, their utility, and illustrative examples* are summarized in Table 2:

Table 2 Common Health Policy Evaluation Types

Evaluation Type	What it Shows	When to Use	Why it is Useful	Examples
Formative Evaluation	Whether the proposed health policy is feasible, appropriate, and acceptable to the target population before it is fully implemented.	During the development of a new health policy. When an existing health policy is being adapted for a different setting. When the focus is on the who, what, when, where, and how.	Allows adjustments to be made before full implementation begins.	Assessing if the health policy can be implemented as planned. Assessing if the health policy will be accepted by the target groups.
Process Evaluation	The extent to which the health policy is being implemented as intended.	Beginning of health policy implementation. During implementation of an existing health policy.	Provides early identification of implementation gaps or challenges. Allows for real-time adjustments to improve policy execution fidelity.	Assessing reach of activities to target groups. Assessing resource mobilization/allocation. Assessing threshold level of participation or exposure to the health policy. Assessing process – for example, procedures, roles, or timeliness for implementing the health policy.
Outcome Evaluation	The degree to which the health policy has achieved its intended outcomes or results. Does not determine causality, only whether outcomes have occurred.	After the health policy has been implemented with the target groups.	Identifies whether the health policy is achieving its stated objectives.	Assessing change in knowledge, attitudes, and behaviors among target groups. Assessing change in policies, regulations, or social norms in target groups. Assessing change in incidence, mortality, and morbidity.

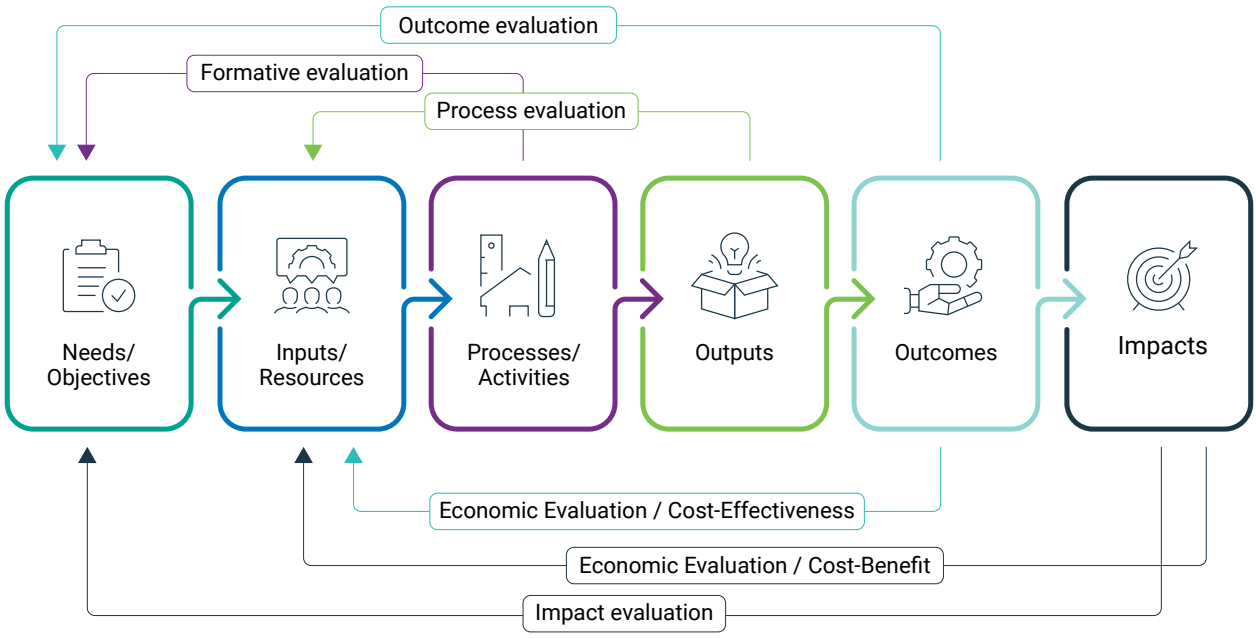
9 See Annex 2 for details.

Evaluation Type	What it Shows	When to Use	Why it is Useful	Examples
Economic Evaluation	Examines health policy effects relative to the costs of the health policy.	At the beginning of health policy implementation. During implementation of an existing health policy. After completion of the health policy intervention.	Helps understand the cost of implementing a health policy and can assess policy effects relative to the cost to produce them.	Assessing if the value of the health policy's outcomes exceeds the cost.
Impact Evaluation	Compares the outcomes of a health policy to estimates of what the outcomes would have been without it. Usually seeks to determine whether activities caused the observed outcomes.	At the end of a health policy. At defined intervals during health policy implementation.	Provides evidence of causal attribution to inform policy and investment decisions.	Assessing the extent to which the outcomes can be related to the health policy as opposed to other external factors-(attribution/causality).

Source: Adapted from CDC's *Program Evaluation Framework Action Guide*, CDC 2024.

Figure 1 visualizes the types of evaluations and their interconnection with the stages of health policy components.

Figure 1 Types of Evaluation and their Interactions with Policy Components[10]



Source: Adapted from *Improving Governance with Policy Evaluation*, OECD 2020.

10 This diagram illustrates key evaluation levels (Input, Process, Output, Outcome, Impact) but is not exhaustive; additional evaluation types and nuances may apply to specific health policies in KSA.

Formative Evaluation

What is Formative Evaluation in Health Policy?

Formative evaluation assesses a health policy during its design or early development phase to determine whether it is feasible, appropriate, and acceptable for the intended target groups and context.[11] Unlike process evaluation, which focuses on how a policy is implemented, formative evaluation is conducted before full-scale implementation begins. It aims to strengthen the policy by identifying design flaws, contextual barriers, and stakeholder concerns to enable refinement before nationwide rollout. This type of evaluation is especially valuable during pilot testing, adaptation to new settings, or early-stage policy development.

Why is Formative Evaluation Important in Health Policy?

Formative evaluation improves the design and contextual fit of a health policy before full-scale implementation. It enables stakeholders to:

1. **Test feasibility**	Determine whether the policy can be implemented given current infrastructure, governance, and workforce capacity.
2. **Assess acceptability**	Evaluate whether the target groups and implementers consider the policy relevant and appropriate.
3. **Identify risks and gaps**	Detect design weaknesses, contextual limitations, or stakeholder concerns in advance.
4. **Support adaptation**	Allow timely refinement to improve alignment with system needs and increase effectiveness.
5. **Increase implementation success**	By resolving issues early, formative evaluation enhances the likelihood of smooth and effective rollout.

Case Study **Formative Evaluation of Integrating Routine Screening for Opioid Use Disorder into Primary Care Settings**

Overview: This case study presents a formative evaluation of routine opioid use disorder (OUD) screening implementation across ten primary care clinics in the United States. Conducted from July 2020 to July 2021, the evaluation took place within a larger multi-site randomized controlled trial (RCT) assessing the effectiveness of the Collaborative Care Model (CoCM) for patients with co-occurring OUD and mental health conditions. While the broader trial used randomization to evaluate clinical outcomes, the formative evaluation was qualitative and non-randomized, aiming to capture real-world implementation experiences, adaptations, and challenges during early rollout.

Policy Context: Despite updated recommendations from the U.S. Preventive Services Task Force supporting universal screening for unhealthy drug use, routine OUD screening remains rare in primary care. Challenges include concerns about staff capacity, stigma, unclear workflows, and lack of comfort addressing positive screens. At the same time, fewer than 21% of individuals diagnosed with OUD receive medications for opioid use disorder (MOUD), underscoring the urgent need to expand access through primary care. This initiative sought to address these issues by integrating OUD screening into CoCM and evaluating its early implementation.

Objectives of the Formative Evaluation: The formative evaluation aimed to:

- Document how population-based OUD screening was implemented across 10 primary care clinics.
- Identify common challenges and contextual barriers encountered during early implementation.
- Explore emerging strategies and adaptations that supported feasibility and workflow integration.
- Inform future implementation support efforts by analyzing real-world experiences and fidelity data.

11 National Academies of Sciences, Engineering, and Medicine. 2023. *Review of four CARA programs and preparing for future evaluations.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26831

Methodology: A qualitative formative evaluation was conducted using a Rapid Assessment Process, guided by the Consolidated Framework for Implementation Research (CFIR). Data Sources included:

- 90 structured observation summaries from meetings between clinics and AIMS Center practice facilitators
- 59 summaries from internal facilitator debriefings
- 10 structured fidelity assessments with clinic teams
- The research team used matrixed data displays and iterative team discussions to identify patterns across clinics, guided by the CFIR's five domains (intervention characteristics, inner setting, outer setting, individual characteristics, and implementation process).

Key Findings: The following cross-cutting barriers were identified:

- Clinics faced uncertainty about who should be screened and at what frequency, especially across different visit types and patient populations.
- Staff and patients found the screening tool difficult to use and interpret, leading to inconsistent implementation.
- Many staff were unsure how to introduce the screening or respond to positive results, especially in the absence of a warm handoff.
- Screening was often missed due to lack of clarity on roles, especially when staff shortages disrupted communication between medical assistants and providers.
- Frequent staff changes and competing demands reduced screening consistency and implementation momentum.
- Few patients screened positive, leading some staff to question the value of the screening effort.
- Internalized and structural stigma toward opioid use reduced patient engagement and staff confidence in addressing substance use.

Promising Implementation Strategies: Through fidelity assessments and observation of implementation meetings, the formative evaluation identified several strategies clinics developed to improve screening feasibility:

- Implementing screening for all patients at every visit to simplify staff decision-making.
- Using patient portals and Electronic Health Records alerts to standardize and track screening.
- Equipping staff with language and confidence to approach patients sensitively.
- Adapting workflows with continuous feedback from facilitators and internal teams.
- Updating clinic policies and publicly signaling that effective medications are available to normalize treatment and reduce patient hesitation.

Conclusion and Implications: Formative evaluation was crucial for identifying real-time barriers and informing adaptive policy strategies during early implementation experiences. The formative evaluation demonstrated that routine OUD screening in primary care is feasible but requires structured support, clear workflows, and continuous adaptation. Key barriers included tool complexity, staff uncertainty, workflow inconsistencies, and stigma. Clinics that employed universal screening, digital tools, and team-based planning saw improved consistency and acceptability.

Source: Austin, Elizabeth J. et al. 2023. "Integrating Routine Screening for Opioid Use Disorder into Primary Care Settings: Experiences from a National Cohort of Clinics." *Journal of General Internal Medicine* 38 (2): 332-340. https://doi.org/10.1007/s11606-022-07675-2

Process Evaluation

What is Process Evaluation in Health Policy?

Process evaluation examines how a health policy is implemented, focusing on its delivery dynamics rather than its outcomes.[12] It analyzes the execution process, the interaction of policy components, and operational strengths or weaknesses such as funding delays or staff training gaps to ensure effective functioning. By offering actionable insights, it refines implementation and aligns policies with their intended design.

Why is Process Evaluation Important in Health Policy?

Process evaluation helps to ensure that policies are implemented as intended. It provides insights that allow stakeholders to:

1. **Assess implementation fidelity**	Ensure that the policy is delivered as planned, preserving its core objectives.
2. **Identify barriers and facilitators**	Understand what supports or hinders implementation, such as resource constraints or stakeholder dynamics.
3. **Enhance policy quality**	Use feedback to improve implementation over time.
4. **Inform decision-making**	Provide real-time data to guide course corrections and effective resource use.

Case Study **Process Evaluation of the Implementation and Delivery of Nurse-Family Partnership in British Columbia, Canada**

Overview: The *Nurse-Family Partnership (NFP)* program, delivered across five regional health authorities in British Columbia (BC), Canada (2013–2018), provided structured home visits by public health nurses to socioeconomically disadvantaged, first-time pregnant and parenting girls and young women. As part of the *British Columbia Healthy Connections Project (BCHCP)*, a comprehensive **process evaluation** was conducted to explore how NFP was implemented, adapted, and experienced in real-world public health systems. This evaluation illustrates how process evaluation can assess fidelity, reach, supervision, acceptability, and contextual dynamics in complex interventions.

Policy Context: Initiated in response to provincial child development and health equity goals, NFP was integrated into BC's public health services and aligned with the Ministries of Health and Children & Family Development. With variation across urban and remote geographies, the policy required tailored delivery by experienced nurses, supported by interministerial and academic collaboration. NFP aimed to improve pregnancy outcomes, child development, and maternal life-course.

Objectives of the Process Evaluation: The evaluation aimed to:

- Assess fidelity to NFP core model elements.
- Measure reach, dose delivered/received, and program participation.
- Explore program acceptability among public health nurses (PHNs), supervisors, and managers.
- Examine implementation barriers, workforce dynamics, and retention.
- Analyze reflective supervision practices and team structures.
- Document adaptations in rural, remote, and high-adversity contexts.
- Describe how nurses supported clients facing intimate partner violence (IPV), mental health, or child protection challenges.

12 Grant, A., Carol Bugge, and Mary Wells. 2020. "Designing process evaluations using case study to explore the context of complex interventions evaluated in trials." *Trials* 21: 982. https://doi.org/10.1186/s13063-020-04880-4.

Methodology: A mixed-methods study (2013–2018) used qualitative and quantitative data:

- **Qualitative:** 343 transcripts from 82 PHNs, 19 supervisors, and 23 decision-makers, including 38 exit interviews.
- **Quantitative:** NFP fidelity reports (visit timing, dose), encounter logs (n=14,000+), supervision records, and session data.

Key Metrics: Home visits completed (78.9%), average visits per client (37), mean visit duration (75–80 min), travel distance (average 14.3 km), and reflective supervision coverage (62.4%).

Key Findings:

- **Fidelity:** NFP was implemented in line with core model elements. Clients received an average of 37 visits; 79% of scheduled visits were completed. Most visits occurred in-home (70%) and lasted 75–80 minutes.
- **Reach Challenges:** Only 27.7% enrolled by 16 weeks gestation, below the 60% benchmark.
- **Supervision:** 62.4% of reflective supervision sessions occurred; average session length was 59 minutes. PHNs emphasized its essential role in clinical growth and emotional resilience.
- **Acceptability & Retention:** Nurses found NFP deeply rewarding, citing full-scope practice, structured education, and strong team cohesion as motivators. Turnover was linked to geography, workload, or job insecurity.
- **Adaptations:** In rural/remote regions, nurses used flexible visit locations, accommodated long travel (up to 120 km), and occasionally used telehealth.
- **Client Needs:** Nurses supported clients with complex adversities—homelessness, IPV, mental illness—requiring tailored strategies and deep therapeutic alliances.
- **Organizational Factors:** Management support, supervision quality, and reflective practice strongly influenced implementation success.

Conclusion and Implications: The BCHCP process evaluation demonstrated that NFP can be delivered with fidelity and adaptability across varied BC settings. It provided critical insight into how implementation structures, nurse experience, and contextual realities shape service quality. This case exemplifies process evaluation's role in strengthening complex public health interventions through real-time learning and system-responsive improvements.

Source: Jack, S. M. et al. 2020. *Implementation and delivery of Nurse-Family Partnership in British Columbia, Canada: A synthesis of selected findings from the British Columbia Healthy Connections Project Process Evaluation (2013–2018)*. https://phnprep.ca/wp-content/uploads/2021/09/BCHCP_Process-Evaluation_Final-Report.pdf.

Outcome evaluation

What is Outcome Evaluation in Health Policy?

Outcome evaluation systematically assesses a health policy's effects to determine if it meets its intended goals.[13] It focuses on measuring changes such as improved health status, healthcare access, or system efficiency resulting from the policy, offering a clear picture of its impact on the target population. This approach is also vital to understand whether a policy delivers its promised benefits.

13 Office for Health Improvement and Disparities (U.K.). 2018. "Outcome evaluation: evaluation in health and wellbeing." GOV.UK. https://www.gov.uk/guidance/evaluation-in-health-and-wellbeing-outcome.

Why is Outcome Evaluation Important?

Outcome evaluation shows whether a policy is achieving its intended results. It provides evidence that helps to:

Ensure accountability	Offer proof of effectiveness, enabling stakeholders—funders, policy makers, and the public—to hold implementers accountable.
Drive improvement	Identify gaps in performance, supporting adjustments to enhance results over time.
Facilitate learning	Uncover what works and why, informing future policy design and knowledge sharing.
Validate impact	Confirm the policy's contribution to intended changes, justifying its continuation or revision.

Case Study **Outcome Evaluation of the Ottawa Model for Smoking Cessation in Ontario, Canada**

Overview: The Ottawa Model for Smoking Cessation (OMSC), implemented in primary care settings across Ontario, Canada, exemplifies outcome evaluation's role in assessing health policy impact. This evidence-based intervention, evaluated in 2007–2009 by the University of Ottawa Heart Institute, uses the 3 A's framework—Ask, Advise, Act—to boost cessation rates. It demonstrates how outcome evaluation measures effectiveness and informs policy refinement.

Policy Context: Rolled out in 32 primary care practices, OMSC addressed Canada's 19% smoking prevalence (2007) by training providers to identify smokers (Ask), counsel quitting (Advise), and offer support or medication (Act). It aimed to reduce smoking-related diseases like lung cancer and heart disease.

Objectives of the Outcome Evaluation: The evaluation aimed to:

- Measure increases in cessation intervention delivery.
- Assess patient quit rate improvements.
- Identify success factors (e.g., training, fidelity).
- Evaluate cost-effectiveness and scalability potential.

Methodology: A before-and-after design included:

- **Data Collection:** Surveyed 3,800 patients and 481 providers across 32 practices at baseline and six months post-implementation.
- **Outcome Measures:** Tracked Ask, Advise, and Act rates, plus patient quit status, via statistical analysis.
- **Fidelity Assessment:** Linked outcomes to adherence to 10 Best Practices.
- **Cost Analysis:** Estimated savings from reduced smoking-related healthcare use.

Key Findings:

- **Intervention Delivery:** Ask rose from 54.9% to 70.8%, Advise from 40.1% to 64.7%, Act from 34.8% to 59.6% (all significant).
- **Quit Rates:** Quit attempts increased by 15%; sustained quit rates rose approximately 10% from baseline – for example, 5% to 15%, estimated.
- **Influencing Factors:** Practices with 8+ Best Practices saw higher quit rates, highlighting fidelity's role.
- **Cost-Effectiveness:** Reduced healthcare use, such as hospitalizations, suggested savings, estimated at $500 per quitter annually.

Conclusion and Implications: OMSC's evaluation confirmed its effectiveness in raising intervention delivery and quit rates, with cost-effective benefits tied to fidelity. It supported scalability, though long-term cessation and disease reduction need further study, underscoring outcome evaluation's role in evidence-based policy.

Source: Papadakis, S. et al. 2016. *Increasing rates of tobacco treatment delivery in primary care practice: Evaluation of the Ottawa Model for Smoking Cessation*. https://doi.org/10.1370/afm.1909.

Impact evaluation

What is Impact Evaluation?

Impact evaluation examines the long-term, causal effects of a health policy on health outcomes, systems, or populations. It goes beyond immediate results—assessed in outcome evaluation—to measure the policy's broader contribution, including intended and unintended changes in societal well-being, equity, or systemic performance. By using rigorous methods to attribute effects to the policy, impact evaluation answers the question, "What difference did this make?" and reveals its transformative impact over time.

Why is Impact Evaluation Important?

Impact evaluation helps to determine a policy's long-term value and system-wide effects. It generates evidence that supports efforts to:

1. **Assess long-term effectiveness**	Evaluate sustained success—beyond short-term outcomes—such as reduced disease burden or improved equity, justifying continued support.
2. **Improve future policy design**	Highlight strengths and weaknesses, guiding refinements for greater impact in future iterations.
3. **Uncover broader effects**	Detect unintended outcomes—positive (e.g., economic gains) or negative (e.g., access disparities)—expanding the policy's impact profile.
4. **Ensure accountability**	Provide evidence of meaningful change, demonstrating transparency to funders, agencies, and the public.
5. **Support informed decisions**	Deliver data-driven insights, including causal attribution, to enable evidence-based resource allocation.

When Should an Impact Evaluation Be Conducted?

Impact evaluations occur post-implementation, once long-term effects emerge, depending on the policy's goals, such as chronic disease reduction, and data availability. Timing balances sufficient impact manifestation with actionable relevance, ensuring results inform current decision-making.

Case Study Impact Evaluation of The Free ART Program in South Africa

Overview: South Africa's free antiretroviral therapy (ART) program, launched in 2004, exemplifies impact evaluation by assessing its long-term effects on HIV/AIDS outcomes nationwide. Targeting Black Africans aged 25–49—the group bearing two-thirds of HIV cases—this initiative expanded treatment access through public facilities. Evaluated using longitudinal survey data collected between 2006 and 2016, the program demonstrates how impact evaluation can quantify sustained health improvements.

Policy Context: Initiated in 2004, the ART program tackled South Africa's HIV epidemic (19% national prevalence, 2004, Stats SA), rolling out free treatment to curb mortality and enhance health. By 2018, 4.6 million of 7.7 million HIV-positive individuals received ART, reversing a crisis that killed over 300,000 annually prerollout.

Objectives of the Impact Evaluation: The evaluation aimed to:

- Measure reductions in HIV-related mortality among Black Africans aged 25–49.
- Assess improvements in self-reported health for this group.
- Estimate the causal impact of ART availability on population health outcomes.

Methodology: A difference-in-differences design used

- **Data collection:** Longitudinal data from the National Income Dynamics Study (NIDS) tracked over 28,000 individuals nationwide (2008–2016/7).
- **Analysis:** Compared communities with varying ART rollout timing, controlling for confounders – for example, age, wealth.

Indicators: Included annual mortality rates and self-reported health scores.

Key Findings:

- **Mortality reduction:** ART availability reduced mortality by 27% among Black Africans aged 25-49 over 2006–2016.
- **Health improvement:** The likelihood of reporting poor health dropped by 36% in this group, reflecting treatment efficacy.
- **Demographic impact:** Within this high-prevalence group, annual mortality fell by 31%, and poor health reports decreased by 47%.

Conclusion and Implications: The evaluation confirmed ART's dramatic impact, sharply reducing mortality and boosting health among South Africa's most affected population. It highlighted the power of nationwide treatment access, offering lessons for global HIV strategies while emphasizing the need for ongoing monitoring of long-term effects.

Source: Burger, C., Ronelle Burger, and Eddy van Doorslaer. 2022. *The Health Impact of Free Access to Antiretroviral Therapy in South Africa.* https://doi.org/10.1016/j.socscimed.2022.114832.

Economic evaluation

What is an Economic Evaluation?

Economic evaluation assesses the costs and health outcomes of a policy to determine its value-for-money. It compares resource use (funding, staff) with benefits achieved (improved health, lives saved), aiding policy makers in understanding how to allocate resources efficiently. Common methods include Cost-Effectiveness Analysis (CEA), and Cost-Benefit Analysis (CBA), each measuring costs against health benefits from different perspectives.[14]

Why is Economic Evaluation Important?

Economic evaluation is essential for making informed health policy decisions by identifying which health policy delivers the best value for money. It helps policy makers:

1. **Assess efficiency**	Compare costs and health outcomes to determine which policies provide the greatest benefit per unit of resource.
2. **Guide resource allocation**	Identify cost-effective strategies—such as vaccination programs or new technologies—especially in resource-constrained settings.
3. **Ensure equity**	Analyze how costs and benefits are distributed across population groups, informing policies that promote health equity.
4. **Provide evidence**	Supply robust data to support the prioritization of health policies that offer the highest health returns for investment.

14 Centers for Disease Control and Prevention. 2024. "Economic evaluation: Overview." POLARIS. https://www.cdc.gov/polaris/php/economics/index.html.

Case Study Economic Evaluation of The HPV Vaccination Program in Australia

Overview: Australia's Human Papillomavirus (HPV) vaccination program, launched in 2007, exemplifies economic evaluation by assessing its costs and health outcomes. Targeting girls aged 12–13, it aimed to reduce cervical cancer through a national rollout. This evaluation, modeled from 2007 onwards, shows how economic evaluation informs efficient health policy.

Policy Context: Introduced in 2007, the HPV program addressed Australia's cervical cancer rate of 7 cases per 100,000 women annually (AIHW, pre-2007). Delivered free via schools, it cost approximately AUD 240 million initially, seeking to lower cancer incidence through widespread vaccination.

Objectives of the Economic Evaluation: The evaluation aimed to

- Assess the cost-effectiveness of HPV vaccination in reducing cervical cancer.
- Compare program costs against projected health benefits.
- Examine resource use efficiency for long-term outcomes.

Methodology: A Cost-Effectiveness Analysis (CEA) was used, involving

- **Data collection:** Combined vaccination costs (AUD 100–150 per dose) with modeled health outcomes (cancer cases averted).
- **Analysis:** A dynamic transmission and Markov model compared vaccination versus no-vaccination scenarios, projecting effects to 2035.
- **Indicators:** Measured costs per Quality-Adjusted Life Year (QALY) gained, discounted at 5%.

Key Findings:

- **Cost-Effectiveness:** The quadrivalent vaccine cost AUD 18,000 per QALY gained, below Australia's AUD 50,000 threshold.
- **Health outcomes:** Projected 50–60% reduction in cervical cancer incidence long-term with continued vaccination and screening.
- **Efficiency:** Modeled savings in treatment costs outweighed vaccination expenses over decades.

Conclusion and implications: The evaluation confirmed the HPV program's value-for-money, providing evidence of substantial health benefits at an acceptable cost. It supports sustained vaccination efforts, offering a model for cost-effective cancer prevention strategies globally.

Source: Simms K. T. et al. 2016. "Cost-effectiveness of the next generation nonavalent human papillomavirus vaccine in the context of primary human papillomavirus screening in Australia: a comparative modelling analysis." *Lancet Public Health* 1 (2): E66–E75. https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(16)30019-6/fulltext.

Note: Outcomes are modeled projections, not empirical 2017 data.

07

How to Conduct Health Policy Evaluations

This section describes the main standards and steps required to conduct health policy evaluations in practice, which ensure to assess policies consistently, produce reliable findings, and improve health outcomes – see Figure 2 below. This section specifies the four core standards and the overview of the six key steps, followed by subsections that provide detailed guidance on applying these steps.

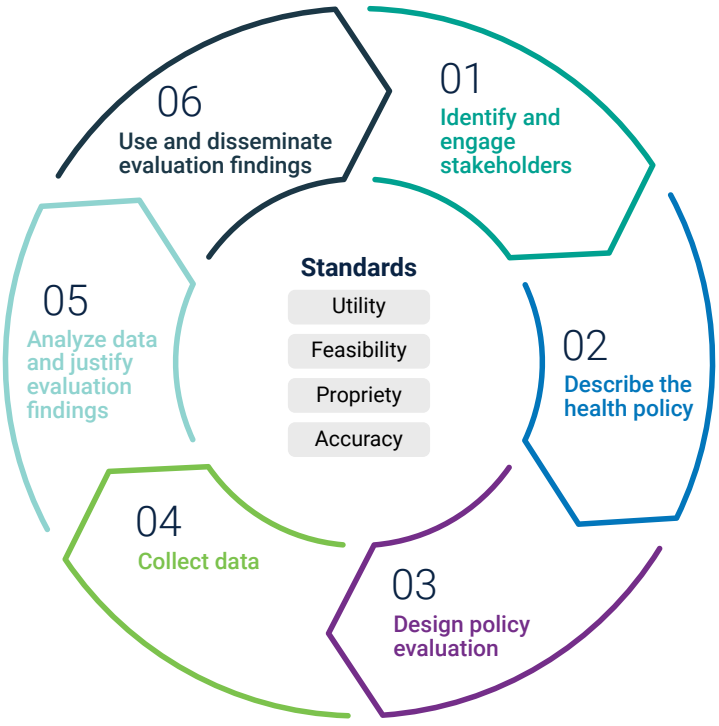
The health policy evaluation conducted based on this guideline follows the *four evaluation standards* which guide the overall evaluation process and ensure its quality:

1. **Utility**	This standard ensures that the evaluation meets the needs of its users, delivers relevant information to those who need the evaluation, and considers the interests of the stakeholders at each step of the evaluation process.
2. **Feasibility**	This standard keeps the evaluation practical within available resources, requires the evaluators to consider the available time, resources, and expertise to complete the evaluation.
3. **Propriety**	This standard maintains fairness and ethical conduct throughout the evaluation process, and ensures respecting the rights and well-being of individuals and stakeholders involved.
4. **Accuracy**	This standard requires the evaluation to provide reliable and precise information.

The steps described in the following subsections offer a clear framework to evaluate a health policy. The health policy evaluation is conducted following six main steps:

1. **Identify and engage stakeholders**	Involve those affected by or interested in the policy.
2. **Describe the health policy**	Define the policy's purpose and expected health outcomes.
3. **Design the policy evaluation**	Plan the methods, scope, and data sources.
4. **Collect data**	Gather credible information to address evaluation questions.
5. **Analyze and justify evaluation findings**	Interpret data to draw evidence-based conclusions.
6. **Use and disseminate the evaluation findings**	Present findings and apply them to enhance the policy.

Figure 2 Overview of the Steps and Standards of Health Policy Evaluation



Source: Adapted from *CDC's Program Evaluation Framework Action Guide*, 2024.

Which Steps to Follow in Conducting Health Policy Evaluations?

Health policy evaluations follow a systematic, six-step process guided by core standards to ensure effective and structured assessment. These steps build upon each other sequentially and are described in detail in the following subsections.

In line with this framework, this guideline presents findings from the evaluation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners in Saudi Arabia, structured according to the six-step evaluation process.

Step 1: Identify and Engage Stakeholders

Why is Stakeholder Identification and Engagement Important?

Involving and committing all key stakeholders at each step is crucial for the success of health policy evaluation. Engaging stakeholders leads to stronger relationships and better communication between the evaluators and the stakeholders, and common understanding of health policies evaluated. That also creates a platform for stakeholders to put any specific questions that may arise during the evaluation process. Engaging stakeholders also facilitates access to necessary data, acknowledgement of the judgments about the evidence gathered, and interpretation of the evaluation findings. Stakeholder engagement also increases the credibility of the analysis and the likelihood that the findings will be used; otherwise, the evaluation could be ignored, criticized, or resisted.

Who are the Potential Stakeholders in a Health Policy Evaluation?

Stakeholders cover a wide range of participants in health policies, including general public, patients, healthcare providers, health system regulators, health financing organizations, patients, academia, private firms, and many others. They are critical players of any evaluation process. Identifying and engaging stakeholders in an evaluation process require transparency and predetermined criteria that align with evaluation standards.[15] The selection of stakeholders is based on their role—to enhance the credibility of the evaluation, ensure effective day-to-day implementation, advocate for necessary changes, and secure funding or authorization for the continuation or expansion of a health policy. Stakeholders in health policy evaluations could be classified under the following groups:

1. **Stakeholders implementing health policy**	Include the stakeholders who have responsibility to develop, approve, or implement the health policy being evaluated and to perform health policy evaluation.
2. **Stakeholders affected by health policy**	Include the stakeholders who are responsible for scrutinizing government decisions and spending as well as the beneficiaries of the health policy being evaluated.
3. **Stakeholders utilizing the evaluation findings**	Include the stakeholders who are responsible for future policies and use evaluation results for further analysis.

Case Study **Mandatory Medical Malpractice Insurance for Other Health Practitioners in KSA**

After the successful launch and implementation of the mandatory medical malpractice insurance for physicians and dentists in 2022, KSA's government decided to expand the coverage of the policy to other health practitioners in 18 specialties. Following the launch of the expansion, an assessment has been conducted to evaluate the awareness and opinion of other health practitioners about the policy.

The stakeholders of the Mandatory Medical Malpractice Insurance for Other Health Practitioners are categorized into the groups described in Table 3.

15 Centers for Disease Control and Prevention (CDC). 2024. *CDC's Program Evaluation Framework Action Guide*, pg. 18. https://www.cdc.gov/evaluation/media/pdfs/2024/12/FINAL-Action-Guide-for-DFE-12182024_1.pdf.

Table 3 Stakeholder Groups under the policy on "Mandatory Medical Malpractice Insurance for Other Health Practitioners" (sample)

Stakeholders	Stakeholder groups					
	Stakeholders responsible for policy-making	**Stakeholders responsible for similar policies in future**	**Stakeholders responsible for evaluation analyses**	**Stakeholders scrutinizing government decisions and spend**	**Participants / recipients of the policy**	**Stakeholders responsible for health policy implementation**
Ministry of Health	☒	☒	☒	☒	☒	☒
Saudi Health Council	☒	☒	☒	☒	☒	☒
Patient Safety Center	☐	☐	☐	☐	☐	☐
Saudi Commission for Health Specialties	☐	☐	☐	☐	☐	☐
Ministry of Health Hospitals	☐	☐	☐	☐	☒	☒
Ministry of Defense	☐	☐	☐	☐	☒	☒
Ministry of Interior	☐	☐	☐	☐	☒	☒
Ministry of National Guard	☐	☐	☐	☐	☒	☒
King Faisal Specialist Hospital and Research Centre	☐	☐	☐	☐	☒	☒
Ministry of Education (Health Affairs)	☐	☐	☐	☐	☒	☒
Patients	☐	☐	☐	☐	☒	☐
Private Sector	☐	☐	☐	☐	☒	☒
Health care Practitioners	☐	☐	☐	☐	☒	☒

Depending on the health policy, its coverage, benefits, and impacts as well as implementation arrangements, stakeholders are also mapped, based on their power and interests, to understand their position and plan the engagement activities.

How to Engage Stakeholders?

After identifying key stakeholders, the roles of the stakeholders in the evaluation process are assessed and a stakeholder engagement plan is developed to ensure their support and input in the subsequent steps of the health policy evaluation. Such a plan clarifies each stakeholder's role and holds all involved parties accountable for the evaluation's success (Table 4), including to increase the credibility of the evaluation, implement health policies, advocate changes as a result of the evaluation findings, and support the expansion or continuation of a policy.

Table 4 Roles of Stakeholders in the Evaluation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (sample)

Stakeholders	Stakeholders' role in policy evaluation process			
	Increase credibility of the evaluation	Implement health policies that are central to this evaluation	Advocate for changes to institutionalize the evaluation findings	Fund/authorize the continuation or expansion of the policy
Ministry of Health	☐	☒	☒	☒
Saudi Health Council	☒	☐	☒	☒
Patient Safety Center	☐	☐	☒	☐
Saudi Commission for Health Specialties	☐	☐	☒	☐
Ministry of Health Hospitals	☐	☒	☒	☐
Ministry of Defense	☐	☒	☒	☐
Ministry of Interior	☐	☒	☒	☐
Ministry of National Guard	☐	☒	☒	☐
King Faisal Specialist Hospital and Research Centre	☐	☒	☒	☐
Ministry of Education (Health Affairs)	☐	☒	☒	☐
Patients	☐	☐	☒	☐
Private Sector	☐	☒	☒	☐
Health care Practitioners	☐	☒	☒	☐

Following the stakeholder mapping and identifying the roles of the individual stakeholders, the evaluation team plans stakeholder engagement in the evaluation process (Table 5). Hence, stakeholders are involved in making their inputs to the evaluation process and take the ownership of the overall evaluation.

Table 5 Stakeholder Involvement Plan in the Evaluation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (Sample)

Stakeholders	Stakeholders' inputs in policy evaluation process				
	Step 2 - Describing health policy	Step 3 - Designing policy evaluation	Step 4 - Data collection	Step 5 - Justify conclusions	Step 6 - Use and disseminate findings
Ministry of Health	☐	☐	☒	☐	☐
Saudi Health Council	☒	☒	☒	☒	☒
Patient Safety Center	☐	☐	☒	☐	☐
Saudi Commission for Health Specialties	☐	☐	☒	☐	☐
Ministry of Health Hospitals	☐	☐	☒	☐	☐
Ministry of Defense	☐	☐	☒	☐	☐
Ministry of Interior	☐	☐	☒	☐	☐
Ministry of National Guard	☐	☐	☒	☐	☐
King Faisal Specialist Hospital and Research Centre	☐	☐	☒	☐	☐
Ministry of Education (Health Affairs)	☐	☐	☒	☐	☐
Patients	☐	☐	☐	☐	☐
Private Sector	☐	☐	☒	☐	☐
Health care Practitioners	☐	☐	☒	☐	☐

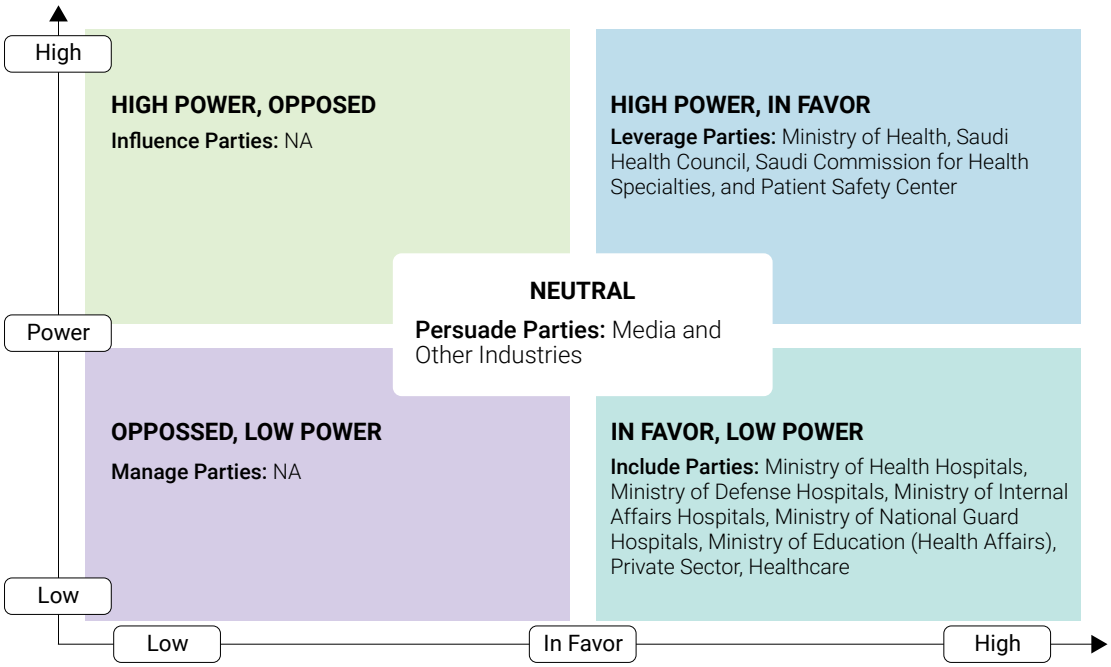
Source: Adapted from Iskarpatyati et al 2011, and *CDC's Program Evaluation Framework Action Guide*, 2024.

In addition to identifying stakeholders' roles and involvement, the evaluation team should develop a targeted engagement strategy based on a stakeholder power-position matrix. Stakeholders can be grouped into five categories—**Influence, Leverage, Manage, Include**, and **Persuade**—depending on their level of power and their stance toward the evaluation. Each group requires a tailored engagement approach:

1. **Influence Parties:** These are high-power actors whose opposition may hinder the evaluation process or the uptake of its findings. Evaluators should engage them early to address concerns about potential bias or the implications of evaluation results. Participatory approaches such as consultative workshops can help build mutual understanding and trust.
2. **Leverage Parties:** These high-power supporters can institutionalize evaluation practices, allocate resources, and promote the use of findings. Evaluators should actively engage them to secure political buy-in, ensure sufficient resources, and facilitate dissemination and application of results.
3. **Manage Parties:** These are low-power opponents who may have concerns or fears about the evaluation but limited ability to obstruct it. Evaluators should manage them through transparency, involvement in the evaluation design, consistent communication, and reassurance that findings will be used constructively, not punitively.
4. **Include Parties:** These low-power supporters can offer valuable insights and help to extend the reach of findings. Evaluators should involve them in data collection, analysis, and validation activities, and support their role as knowledge brokers in dissemination efforts.
5. **Persuade Parties:** These neutral or unengaged actors can become valuable allies if adequately informed. Evaluators should raise awareness of the evaluation's objectives, tailor messaging to their interests, and engage them in learning and dissemination events to encourage support.

This engagement typology helps to ensure that stakeholder participation is strategic, inclusive, and aligned with the overall evaluation goals. The matrix in Figure 3 provides a template to operationalize this typology by classifying stakeholders and identifying tailored engagement strategies.

Figure 3 Stakeholder mapping for the policy on "Mandatory Medical Malpractice Insurance for Other Health Practitioners"[16] (sample)



Source: Developed by the authors using evaluation findings from the implementation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (2024), in consultation with the Saudi Health Council.

16 The evaluation findings of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners, introduced by the Kingdom of Saudi Arabia (KSA), serve as the sample case presented in this guideline.

Checklist for Step 1: Identify and Engage Stakeholders	
☐	Identify and categorize stakeholders into three main groups: those impacted by the policy, those involved in its implementation, and those who will use the evaluation findings.
☐	Review the stakeholder list to pinpoint key stakeholders that are crucial for enhancing the evaluation's credibility, ensuring policy implementation, advocating for institutionalization, or approving the continuation or expansion of the evaluation.
☐	Engage individual stakeholders or representatives from relevant organizations, understand their interest and powers.
☐	Develop a plan for stakeholder involvement and specify areas where their input is required.
☐	Ensure selected stakeholders regularly participate in crucial steps, such as describing the policy, developing and choosing evaluation questions, and disseminating the evaluation results.

Step 2: Describe the Health Policy

Understanding the Health Policy and Its Key Elements

The next step is to clearly define the health policy being evaluated. This involves thoroughly analyzing the policy, exploring its underlying assumptions, and identifying the key questions that the evaluation aims to answer. Therefore, a comprehensive understanding of the policy is vital, because it informs the subsequent steps of the evaluation, including the development of the evaluation design, the selection of suitable methods, and the identification of relevant indicators and data sources.[17] Table 6 presents elements to examine and detail at this phase of the evaluation.[18]

Table 6 Key Health Policy Components

#	Policy components	Description	Sample from Mandatory Medical Malpractice Insurance for Other Health Practitioners' Policy Evaluation Findings (2024)
1	**Need for policy**	The need for the health policy is defined by the public health issue or other critical challenge(s) that the policy aims to address. This need should be articulated in terms of its consequences for the population or a group, the overall scale of the issue and its prevalence, and significant changes or trends in its incidence or prevalence. Clearly defining the need for the policy is essential for evaluating its relevance and effectiveness in addressing the identified public health challenges.	The significant increase in the number of medical malpractice lawsuits from 1,097 in 2016 to 1,379 in 2018 spurred action to enhance patient safety, minimize medical errors within the healthcare system, and reduce the financial burden on health practitioners in KSA. The increasing number of lawsuits signaled challenges in quality of care and risks of financial burdens for health practitioners. Therefore, the KSA government decided to extend the existing Mandatory Medical Malpractice Insurance policy to include other specified healthcare providers across 18 specialties.
2	**Target Groups**	Target groups are the specific populations or stakeholders that the health policy seeks to engage in addressing the public health issue. Understanding and clearly defining the target groups is crucial for evaluating how well the health policy is reaching the intended audience and whether it is effectively mobilizing these groups toward achieving the desired health outcomes.	The main target group of the policy is the health practitioners from the following 18 specialties: nurses, pharmacists, anesthesia specialists, midwifery specialists, diagnostic radiologists, diagnostic radiology technicians, emergency medical service providers, lab specialists, physiotherapists, speech and language pathologists, respiratory therapists, nutrition specialists, audiologists, phlebotomy specialists, ophthalmologists, operating room technicians, cardiac perfusion specialists and blood draw optometry specialists.

17 Note: In cases where evaluators have limited prior knowledge or documentation of the policy, it may be necessary to begin with a detailed description of the policy itself. In such instances, this step might precede stakeholder engagement to ensure clarity and alignment throughout the evaluation process.

18 *CDC's Program Evaluation Framework Action Guide*. https://www.cdc.gov/evaluation/media/pdfs/2024/12/FINAL-Action-Guide-for-DFE-12182024_1.pdf.

3	**Activities**	Activities are the actions undertaken by the health policy and its implementing bodies to achieve the desired outcomes within the target population. Evaluating these activities involves assessing whether they are being implemented as planned and whether they are contributing to the intended outcomes.	The main activities under the policy include the dissemination of the policy, standardization of the health insurance guidelines, and improvement of the awareness among health practitioners about the medical malpractice insurance.
4	**Resources / Inputs**	Resources and inputs refer to the people, funding, and information required to implement health policy activities. In the evaluation context, it is important to assess whether the necessary resources are available. If the intended outcomes are not being achieved, the evaluation should consider whether resource constraints or misallocations are contributing factors.	The main resources used for the policy implementation are the human resources.
5	**Outputs**	Outputs are the direct products of the health policy activities, typically represented as tangible deliverables. Evaluating outputs involves measuring whether the activities are producing the expected results in quantifiable terms.	An increasing number of insured healthcare practitioners, improved knowledge on malpractice insurance, and establishment of clear insurance guidelines are the main expected outputs of the policy.
6	**Outcomes**	Outcomes represent the actual changes that occur as a result of the health policy, directly related to its goals and objectives. Evaluating outcomes involves assessing short-term and long-term changes: • Short-term outcomes should be evaluated to determine whether the policy is beginning to have the desired effects, such as changes in attitudes, behavior, or health status, – for example, within the first one to three years of implementation. • Long-term outcomes involve evaluating the sustained effects of the policy such as broader changes in health behavior, practices, or status – for example, over four to six years – that build on short-term outcomes.	Improved patient safety, reduced incidence of medical errors, enhanced trust in healthcare services, and reduced financial burden on the health practitioners are the main expected outcomes of the Medical Malpractice Insurance for Other Health Practitioners' Policy.
7	**Impact**	Impact refers to the broader, after long-term results of health policy implementation. Evaluating impact involves assessing the overall effect of the policy on public health, including organizational-, community-, or system-level changes. This evaluation focuses on whether the policy has led to improved public health conditions, enhanced system capacity, and significant changes in the policy environment, ultimately achieving its long-term goals.	Improved patient satisfaction, increased life expectancy, better overall healthcare quality, and strengthened healthcare system are the main expected impacts of the Medical Malpractice Insurance for Other Health Practitioners' Policy.

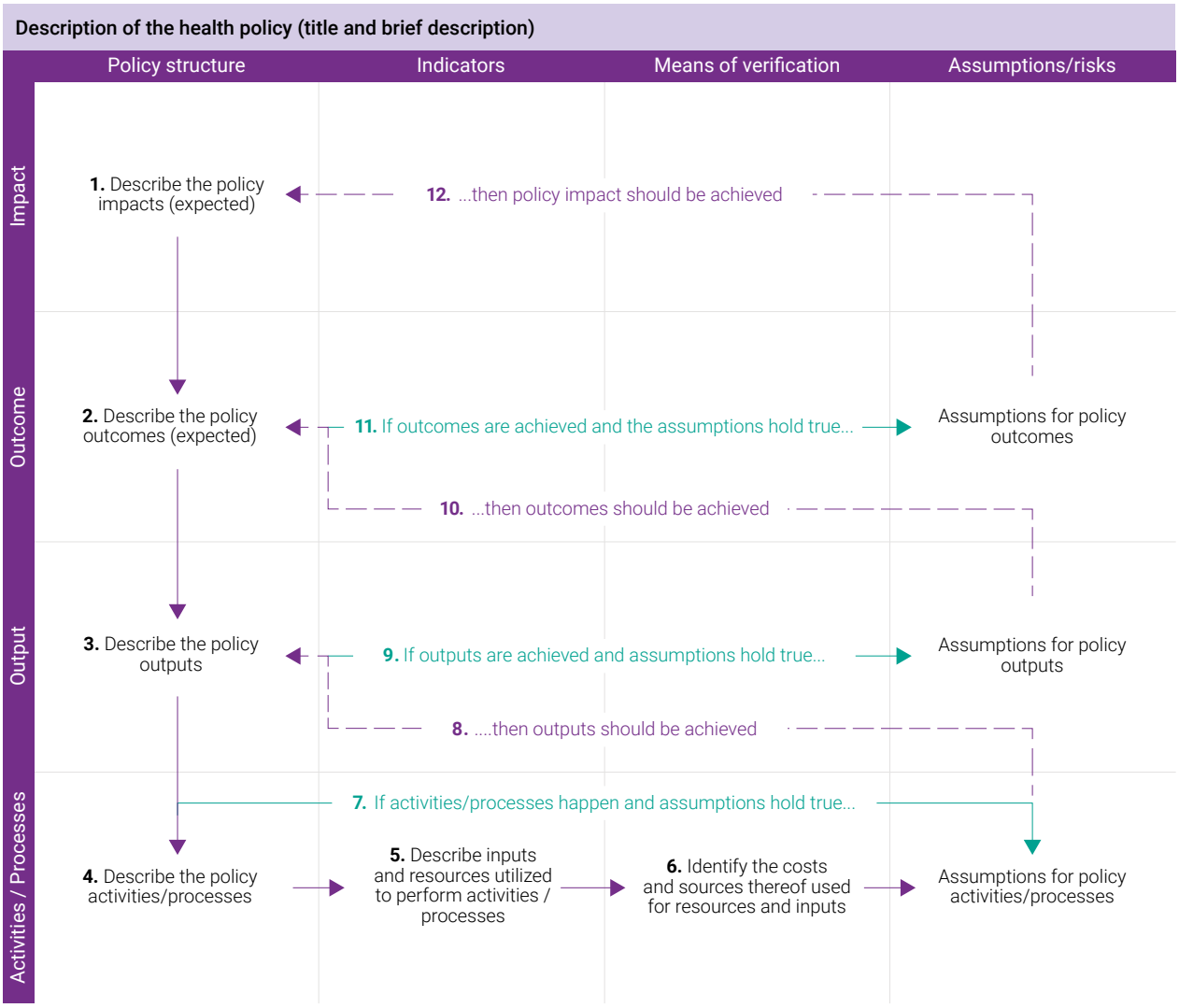
Source: Adapted from *CDC's Program Evaluation Framework Action Guide*, 2024, and the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners, 2024.

Beyond understanding the individual components of a health policy, it is crucial to examine how they interact with one another. These interrelationships can be effectively explored using a logic model which serves as a systematic approach to organizing and illustrating the relationships between planned activities and their measurable objectives, offering a detailed explanation of health policy.

The logical framework presented in this guideline helps evaluators clearly outline the key components of a policy and their logical interconnections. By mapping outputs, activities, and underlying assumptions, the framework demonstrates a clear progression from inputs and actions to the desired outcomes and impacts. It reflects the underlying Theory of Change of the health policy, typically developed during the policy design stage.[19]

The sample logical framework template below presents the core elements of the framework and illustrates the logical connections among policy components.

Figure 4 Logical Framework Template



Source: Adapted from *Log Framework Handbook*, World Bank.[20]

The logical framework presented above is designed to support a structured evaluation of health policy by applying two complementary approaches: **a vertical (top-down) analysis of policy goals, and a horizontal analysis of measurement components**. Together, these approaches help evaluators to understand *what* a policy aims to achieve and *how* to verify whether those goals are being met.

19 *The Health Policy Maker's Manual: Integrating Data and Evidence* (2024) provides more details on the Theory of Change in policy development.
20 World Bank. 2005. *The Logical Framework (Logframe) Handbook: A Logical Framework Approach to Project Cycle Management*. https://documents1.worldbank.org/curated/en/783001468134383368/pdf/31240b0LFhandbook.pdf.

1. Top-Down Approach: Evaluating the Hierarchy of Policy Goals

The vertical logic of the framework follows a top-down approach, enabling evaluators to trace how each level of the policy contributes to achieving the overarching goals. This mirrors the evaluation process, which typically begins by assessing whether the intended impact has been achieved, and then works backward through outcomes, outputs, and activities.

1.1 **Start at the top (impacts):** The policy's long-term objectives (impacts) are the ultimate changes or benefits the policy aims to achieve.

1.2 **Move to outcomes:** To understand whether the policy is moving towards its impact, evaluators assess the outcomes—the medium-term effects or changes directly resulting from the policy's outputs.

1.3 **Assess outputs:** The next step is to look at the outputs, which are the immediate results of the policy's activities, such as training sessions completed, facilities built, or services delivered.

1.4 **Evaluate activities and processes:** Finally, evaluators examine the activities and inputs—the specific actions and resources used to implement the policy.

This top-down sequence enables evaluators to trace the logic of the policy from its highest-level goals (impacts) to the specific actions taken (activities), determining whether the policy's intended objectives have been met at every level.

2. Horizontal Approach: Measuring and Verifying Success at Each Level

Once the vertical structure is established, the framework applies a horizontal logic to identify how success will be measured and verified across all levels. This includes three key components:

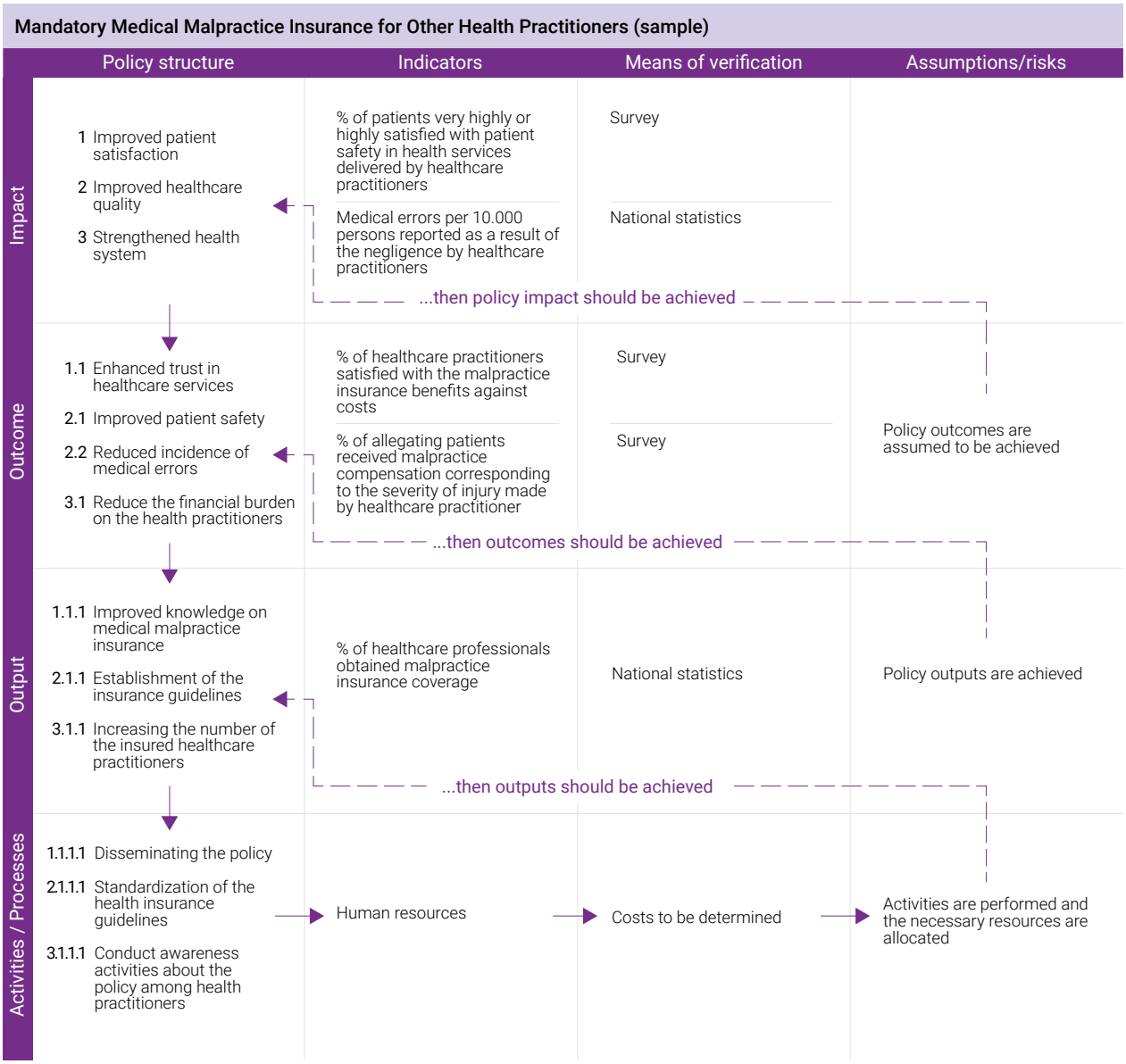
2.1 **Indicators:** These define the metrics or criteria used to measure success at each level, from activities to impacts.

2.2 **Means of verification:** This explains how the data will be collected or verified to ensure the indicators are accurate.

2.3 **Assumptions/risks:** These outline any external factors that may influence the success of each level, helping evaluators to anticipate challenges and risks.

The following sample, based on Saudi Arabia's Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners, demonstrates how the logical framework can be applied to visualize the interconnections among policy components (Figure 5).

Figure 5 Illustrative Logical Framework Developed from the Evaluation Findings of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners



Source: Developed by the authors using evaluation findings from the implementation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (2024), in consultation with the Saudi Health Council.

When constructing a comprehensive logic model, understanding the context—such as the broader environment in which the policy operates, which may present both opportunities and challenges—is another critical consideration. This context may include political dynamics, funding, interagency cooperation, competing organizations and interests, social and economic conditions, and the history of the program or agency, or past collaborations.

Checklist for Step 2: Describe the Health Policy	
☐	Identify the key policy components, including, inputs, activities, outputs, outcomes, and impacts.
☐	Describe each policy component to explore the policy evaluated.
☐	Develop a logical map of the health policy components and determine their relationships.
☐	Create a visual representation of the logical map.
☐	Determine the current stage of policy implementation.
☐	Analyze the context (political, social, financial, etc.) in which the policy is being implemented.

Step 3: Design Evaluation

What types of methods are used to conduct health policy evaluation?

In practice, most evaluation designs employ a combination of methods, integrating both qualitative and quantitative approaches to address questions related to impact, process, and economic (value-for-money) evaluations.[21] The common policy evaluation methods are categorized into the following groups:

1. Commonly used research methods.
2. Theory-based evaluation methods.
3. Experimental and quasi-experimental evaluation methods.
4. Value-for-money methods.
5. Synthesis methods.

The following subsections discuss in detail the various types of these methods and the considerations for their application.

Commonly Used Research Methods

When evaluating health policies, there are cases where the mechanism by which a policy induces change is so straightforward that its impact can be directly observed or assessed through process evaluation without the need to account for other influencing factors. For example, in a country where a new nationwide vaccination policy is introduced to combat a specific infectious disease, if the disease incidence sharply declines following the implementation of the policy, it can be directly attributed to the vaccination efforts. This scenario reflects clear causality, as the vaccination policy is the primary intervention, and there is strong confidence that no other significant changes would have occurred in the absence of the policy.

Table 7 outlines specific methods that can be applied in such scenarios.[22]

Table 7 Commonly Used Research Methods

Evaluation Methods	Description	Analytical focus	Strengths	Limitations
Interviews and Focus Groups	Interviews are qualitative data collection methods that involve direct, in-depth conversations with individuals to explore their knowledge, experiences, or perceptions about a policy. Focus groups bring together multiple participants to discuss a topic, promoting interaction and a range of views. Both methods are used to gather nuanced, context-rich insights that may not emerge through quantitative approaches.[23]	Interviews enable in-depth exploration of health policies with participants such as policy makers, healthcare providers, and community members. Focus groups are useful for eliciting views from a diverse group of stakeholders, providing a broad perspective on the health policy being evaluated. This method is often used to supplement quantitative data by revealing the rationale behind observed trends.	Can be used to gather detailed insights from individuals directly involved in or affected by the health policies. Helps to uncover underlying reasons behind stakeholders' views and sheds light on patterns emerging from other pieces of evidence, such as quantitative monitoring data. Capture diverse viewpoints and support triangulation.	Can be resource-intensive and time-consuming to conduct and analyze. Does not provide numerical estimates. There may be a risk of bias in the views collected, potentially affecting the validity of the findings.

21 *Magenta Book*, page 41.
22 *Magenta Book*, page 42.
23 Patton, M.Q. 2002. *Qualitative Research and Evaluation Methods*. SAGE Publications.

Evaluation Methods	Description	Analytical focus	Strengths	Limitations
Case Studies	Case studies are a method of inquiry that involves an in-depth examination of a single or small number of cases within their real-life contexts. They are particularly valuable in understanding the implementation and effects of complex health policies.[24]	In-depth investigation of specific health policies within their real-world context. Subjects are often purposely selected to represent unique or typical cases, revealing critical information about the implementation and outcomes of health-related actions. Often uses multiple sources of evidence, such as interviews, documents, and observations to gain a holistic understanding of policy processes and outcomes.	Captures real-life situations in depth and detail, aiding in the understanding of complex health policy issues. Works well in combination with or supplementing other methods, such as surveys. Helps to communicate effective health policies to stakeholders.	It is challenging to generalize findings to different contexts, situations, or health policies, limiting the broader applicability of the insights gained from the case study. Interpretation may be influenced by researcher bias. Often time and resource-intensive.
Surveys	Surveys use structured questionnaires to collect data from a large population or sample.[25]	Commonly used to collect data from a large number of individuals, such as healthcare providers, patients, or members of the public affected by a health policy. Quantifies attitudes, knowledge, and behaviors. Provides population-level data for monitoring or comparison across demographic groups.	An effective method for obtaining information from a large number of participants, providing a broad overview of the health policy being evaluated. Provides data suitable for statistical analysis that, if well-designed, can be generalized to the population of interest.	Less useful for providing in-depth insights into the nuances of health policy. Response-rate issues can decrease the quality and reliability of the findings, potentially leading to biased or incomplete data.

24 Yin, R.K. 2014. *Case Study Research: Design and Methods*. SAGE Publications.
25 Dillman, D.A., Smyth, J.D., and Christian, L.M. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Wiley.

Evaluation Methods	Description	Analytical focus	Strengths	Limitations
Observational Studies (including Ethnography)	Observational studies are qualitative or mixed-methods approaches that involve systematically watching and recording behaviors, interactions, or events as they naturally occur in real-life settings, without interference by the researcher. In the context of health policy evaluation, these studies are used to understand how policies are implemented and experienced in practice, especially by frontline staff and affected populations. Ethnography, a more immersive form of observational research, involves extended engagement in the field—clinics, community settings, or policy offices—where the researcher participates in and observes daily routines to understand the cultural, organizational, and social dynamics influencing policy outcomes. Ethnography may also involve informal conversations, field notes, and reflexive analysis to generate deep, contextualized insights. This method is particularly useful when evaluating complex interventions, uncovering discrepancies between policy as planned and policy as enacted, and identifying unanticipated outcomes.[26]	Involves observing and noting the behavior of participants, including healthcare providers and policy implementers, within their usual environments to understand how health policies impact day-to-day practices. Often supplemented with interviews to contextualize the observations and build relevant theories.	Allows for a deeper understanding of how individuals experience a health policy in practice. Observation can help improve the accuracy of other data by reducing biases that arise from self-reporting.	Participants may alter their behavior if they know they are being observed (the 'Hawthorne effect'), affecting data accuracy. Resource-intensive, with potential ethical implications and practical barriers to implementation.

Source: Adapted from *Magenta Book*, pages 42–43.

26 Hammersley, M., and P. Atkinson. 2019. *Ethnography: Principles in Practice*. Routledge.

Sample Case Evaluating the Implementation of the Mandatory Medical Malpractice Insurance for Other Health Practitioners' Policy

Policy context

Following a three-year implementation of the Mandatory Medical Malpractice Insurance for Physicians, the government of KSA extended the coverage of the policy across 18 specialties of health practitioners. The Saudi Health Council conducted a survey among health practitioners to evaluate the implementation status of the policy.

Focus of the survey

The survey aimed to provide a comprehensive assessment of this policy among other health practitioners by measuring the compliance of healthcare providing entities, the number of insured other healthcare workers, and the level of awareness and understanding of the policy across the health practitioners. The survey objective includes the following:

- Measuring the percentage of healthcare practitioners who possess the mandatory medical malpractice insurance coverage.
- Measuring the level of awareness and understanding of the policy across the healthcare practitioner through a survey.
- Assessing the compliance of public and private entities providing healthcare in requiring their healthcare professionals to be insured against malpractice.
- Identify the challenges and limitations in implementing the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy.
- Provide actionable recommendations to enhance the designing of the implementation process for the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy.

Survey methodology

A cross-sectional study was conducted using a random sample, with two specific questionnaires created—one for nonphysician healthcare practitioners and the other for healthcare facilities providing the service. This design aimed to collect comprehensive data within a specified period from June to September 2024, helping to understand the perspectives of healthcare practitioners and facilities towards the mandatory cooperative insurance policy against medical errors for nonphysician healthcare practitioners.

The study includes several subobjectives: measuring the percentage of non-physician healthcare practitioners who have medical error insurance; assessing the level of awareness and understanding of the policy among nonphysician healthcare practitioners through a questionnaire; evaluating the compliance of public and private healthcare institutions and facilities with the mandatory cooperative insurance requirements for nonphysician healthcare practitioners against medical errors; and identifying challenges in implementing the mandatory medical insurance policy against errors for nonphysician healthcare practitioners.

The study also contributed practical recommendations to enhance the design and implementation of the mandatory medical insurance policy against errors for nonphysician healthcare practitioners, providing valuable information for decision-makers. Additionally, data were analyzed using advanced statistical methods with the IBM® SPSS statistical software to ensure the accuracy and objectivity of the results. Various data analysis methods were used, including descriptive analysis that covered basic statistics such as means, frequencies, and percentages to understand the general distribution of the data. Inferential analysis was also conducted through hypothesis tests such as the T-test and chi-square test to determine if there were statistically significant differences between groups.

Survey sample

A total of 261 responses were collected from the healthcare facilities, which included a variety of institutions such as hospitals, clinics, primary care centers, polyclinics, and optical shops. Additionally, 1,112 healthcare practitioners responded to the questionnaire. After excluding physicians and dental assistants, the sample consisted of 1,090 healthcare practitioners. Following the removal of duplicate responses, the final sample size was 1,069 healthcare practitioners.

Source: Developed by the authors using evaluation findings from the implementation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (2024), in consultation with the Saudi Health Council.

Theory-based evaluation method

Theory-based approaches offer a valuable method for evaluating health policies by focusing on the specific pathways through which a policy is expected to produce its effects.[27] These approaches go beyond simply measuring outcomes; they aim to uncover the underlying mechanisms and contextual factors that drive change. Unlike methods that prioritize precise estimates of effect size, theory-based evaluations emphasize understanding the policy's contribution to observed outcomes, making them particularly useful in complex, real-world settings where multiple factors influence results.

These methods are particularly effective when evaluating complex health policies or those in complex policy environments. In such cases, where measuring exact impact is difficult, theory-based methods help to confirm whether the policy is moving in the intended direction. They provide insights into why a policy worked or did not work, making them valuable for adapting policies to different populations, settings, or time periods.

While theory-based methods can incorporate various evaluation techniques, they are especially suited for those that explore the processes within the policy's implementation. Theory-based impact evaluation methods are particularly appropriate when one or more of the following conditions apply:[28]

- The policy environment is complex, involving a mix of different policies.
- The policy aims to create change within a complex system or in situations where there is adaptive management or modification of the policy.
- Outcomes are emergent and cannot be predicted from the start.
- It is impossible to develop a suitable counterfactual.
- There is a need to understand if the same outcomes would occur in a different setting or context.

The methods outlined below are closely aligned with this approach, offering a framework to investigate the effectiveness of health policies in their specific contexts. Table 8 provides an overview of the most common theory-based methods, their application, and challenges.[29]

Table 8 Types of Theory-Based Evaluation Methods

Evaluation Methods	Description	Analytical Focus	Strengths	Limitations
Realist Evaluation	Investigates how and why a policy works (or fails), for whom, and under what circumstances. It emphasizes identifying underlying mechanisms and the contextual conditions that activate them, offering nuanced insight into complex policies.[30]	Identifies and tests causal mechanisms within different contexts. Emphasizes understanding how stakeholders respond to policy under specific circumstances.	Helps to refine theoretical understanding specific to health policies and informs impact evaluations even when creating a counterfactual is challenging.	Requires significant time, resources, and expertise. The complexity of the analysis can make results hard to interpret or communicate. Rarely provides quantitative effect sizes.
Contribution Analysis	A structured approach to evaluating the contribution a policy or intervention makes to observed outcomes by linking activities and results, while ruling out alternative explanations.[31]	Tests causal claims by developing and verifying a contribution story. Uses Theory of Change as backbone.	Strengthens the argument about the policy's impact through logical reasoning, particularly when direct attribution is difficult. Enhances credibility of causal claims without requiring counterfactuals.	Less effective in situations where there is significant variation in how the policy is implemented or in its outcomes. May oversimplify complex causal pathways.

27 *Magenta Book*, page 37.
28 *Magenta Book*, page 37.
29 *Magenta Book*, page 45.
30 Pawson, R., and Tilley, N. 1997. *Realistic Evaluation*. SAGE Publications.
31 Mayne, J. 2001. "Addressing attribution through contribution analysis: Using performance measures sensibly." *Canadian Journal of Program Evaluation* 16 (1): 1–24.

Evaluation Methods	Description	Analytical Focus	Strengths	Limitations
Process Tracing	A qualitative method used to test hypotheses about causal mechanisms in a specific case. It involves systematically collecting and analyzing evidence to trace how an outcome came about, and whether the sequence of events matches the theory of change or alternative causal pathways.[32]	Examines a single case of policy implementation to test whether the expected causal mechanisms actually lead to the observed outcomes, based on the logic map of the policy.	Can validate causal assumptions after the fact, provided the method is applied rigorously. Allows for the consideration of alternative explanations.	Evidence may not be available or reliable. Time-consuming. May not generalize beyond a single case.
Bayesian Updating	A formal reasoning method that uses probabilistic logic to revise the level of belief in a causal relationship as new evidence becomes available. In policy evaluation, it is often combined with qualitative methods to assess the strength of contribution claims under uncertainty.[33]	Enhances the rigor of health policy evaluations by combining with other methods to update the probability of contribution claims based on new evidence.	Strengthens evidence assessment by using probabilistic reasoning, especially useful when other logical methods are applied.	Resource-intensive, requiring skilled facilitators and extensive expertise.
Contribution Tracing	A structured approach that merges the logic of process tracing with Bayesian reasoning. It engages stakeholders in assessing contribution claims by defining expected outcomes, testing evidence against competing explanations, and quantifying confidence levels in the results.[34]	A mixed-method approach that combines stakeholder participation with systematic criteria for data collection. It includes a "contribution trial" where all stakeholders evaluate what evidence will support or refute a policy's impact.	Efficiently focuses on evidence that can enhance confidence in a policy's contribution, while minimizing bias through critical peer review.	Requires time for the policy's effects to become evident. Must also explore other potential causes. Not suitable for comparing different policies.
Qualitative Comparative Analysis (QCA)	A cross-case analysis method based on set theory that identifies combinations of causal conditions associated with an outcome. It is particularly suited for analyzing complex, context-dependent policies by revealing multiple pathways to success or failure across several cases.[35]	Compares multiple policy cases to identify patterns and combinations of factors associated with desired or undesired outcomes, using qualitative insights.	Well-suited to complexity. Reveals multiple pathways to success or failure. Supports cross-case learning.	Requires consistent data across cases and careful analysis to determine which factors are most successful in different contexts. Interpretation of results can be complex.

32 Beach, D., & Pedersen, R. B. 2013. *Process-Tracing Methods: Foundations and Guidelines*. University of Michigan Press.
33 Befani, B., and Stedman-Bryce, G. 2017. "Process Tracing and Bayesian Updating for Impact Evaluation." *Evaluation* 23 (1): 42–60. https://doi.org/10.1177/1356389016654584.
34 Mayne, J. 2012. "Making causal claims." ILAC Brief 26. https://hdl.handle.net/10568/70211. See also Befani, B., D'Errico, S., Booker, F., and Giuliani, A. 2016. "Clearing the fog: new tools for improving the credibility of impact claims," https://www.iied.org/17359iied.
35 Ragin, C. C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press; and Schneider, C. Q., and Wagemann, C. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge University Press.

Evaluation Methods	Description	Analytical Focus	Strengths	Limitations
Outcome Harvesting	A participatory and retrospective method that identifies outcomes (intended or unintended) and works backward to determine how a policy or program may have influenced them. It does not rely on preset indicators and is especially useful in dynamic or uncertain environments.[36]	Gathers evidence of change and traces it back to assess the health policy's contribution, with ongoing stakeholder involvement for monitoring.	Encourages stakeholder participation, providing real-time insights and clarity during policy implementation. Flexible and adaptive to complex environments. Useful when outcomes are not predefined.	Resource-intensive and may be difficult to apply in large-scale or highly complex health policies.
Most Significant Change (MSC)	A qualitative, participatory technique that gathers significant change stories from stakeholders to understand the most valued outcomes of the policy. It involves group dialogue and selection of stories that reflect meaningful impacts, helping surface unintended or less tangible results.[37]	A participatory method that involves collecting and selecting significant change stories from stakeholders.	Useful for evaluating unpredictable outcomes or when stakeholders need to agree on prioritized results. Fosters understanding and engagement among stakeholders and enhances learning. Captures rich qualitative data.	Time-consuming, resource-demanding, and requires skilled facilitation to ensure meaningful results. Limited generalizability.

Source: Adapted from *Magenta Book*, page 45.

Experimental and quasi-experimental evaluation methods

Experimental and quasi-experimental methods provide robust frameworks for measuring the impact of health policies by comparing observed outcomes in both controlled and real-world settings. The fundamental principle behind the experimental and quasi-experimental methods is the use of a counterfactual—a comparison between observed outcomes in a group that received the health policy intervention and a control group that did not.

- **Experimental designs:** The intervention and control groups are made effectively identical through randomization, ensuring that any differences in outcomes can be attributed to the policy itself – for example, Randomized Controlled Trials (RCTs).
- **Quasi-experimental designs:** The groups may differ in known ways, but these differences are accounted for analytically during the evaluation process.

Depending on the scale of the health policy evaluated, the groups in these evaluations can consist of individuals receiving health care services, healthcare facilities, communities, or even entire regions. It is crucial to minimize interaction or "mixing" between the groups to prevent bias, such as the intervention's "contamination" of control groups.

By collecting and analyzing comparable data from both the intervention and control groups, evaluators should be able to confidently determine the extent to which any observed changes in health outcomes are attributable to the policy. These methods are particularly useful when it is important to quantify the average impact or net benefit of a health policy intervention.

The selection of either an experimental or quasi-experimental evaluation method depends on several key factors:

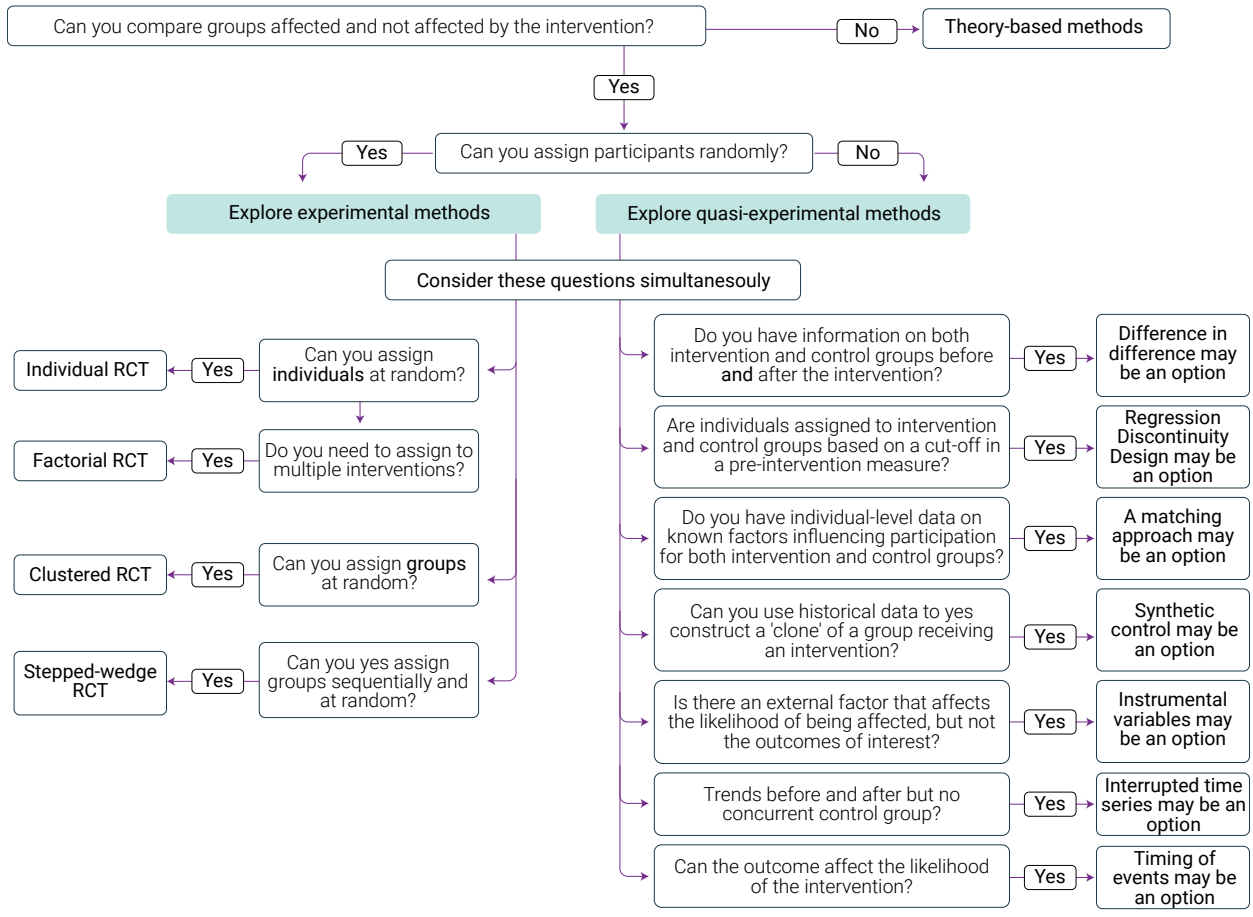
- **Randomization feasibility**. For experimental methods like RCTs,[38] it is crucial to determine if people receiving health care services can be randomly assigned to receive the health policy intervention. If randomization is not feasible, quasi-experimental methods, which do not rely on random assignment, should be considered.

36 Wilson-Grau, R., &and Britt, H. 2012. *Outcome Harvesting*. Ford Foundation.
37 Davies, R., & Dart, J. 2005. *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. https://www.mande.co.uk/wp-content/uploads/2005/MSCGuide.pdf.
38 There are several variations of RCTs to suit different needs in health policy evaluation: (1) **Factorial RCTs** independently randomize participants to multiple interventions; (2) **Cluster RCTs** randomize groups of participants, such as entire communities or healthcare facilities, rather than individuals; (3) **Stepped-Wedge RCTs** apply an intervention sequentially and at random to different groups of participants.

- **Expected effect size**. Consider the anticipated magnitude of the policy's impact on health outcomes. Larger expected effects might be easier to detect with quasi-experimental methods, whereas smaller effects might require the rigor of an RCT to be confidently identified.
- **Data availability**. Evaluate the type and amount of data available for analysis. Experimental methods may require baseline data and follow-up data for both intervention and control groups, whereas quasi-experimental methods may need historical data or data on specific variables that should be used to create a comparable control group.
- **Control group availability**. Determine whether suitable control groups can be identified. For experimental methods, this involves creating control groups through randomization. For quasi-experimental methods, control groups should be identified through natural experiments, historical comparisons, or matching techniques.

To assist in selecting the most appropriate experimental or quasi-experimental method for evaluating a health policy, a set of guiding questions should be used. Figure 6 illustrates the decision-making process for choosing these methods. By following this flowchart, evaluators should systematically assess the conditions under which each method is appropriate, ensuring that the selected approach is well-suited to the specific evaluation context. If none of these methods are suitable, theory-based methods should be considered as an alternative approach.

Figure 6 Choosing Experimental and Quasi-Experimental Methods



Source: Magenta Book, page 47.

Building on this decision-making process, Table 9 provides a comprehensive comparison of the experimental and quasi-experimental methods commonly used in health policy evaluation. This table is designed to assist evaluators by detailing the key characteristics, strengths, and limitations of each method. By consulting this table, evaluators can make informed decisions that align with the specific objectives and context of their health policy assessments.

Table 9 Detailed Overview of Experimental and Quasi-Experimental Methods

Evaluation Method	Description	Analytical Focus	Strengths	Limitations
Randomized Controlled Trial (RCT)	Randomly assigns individuals, communities, or institutions to treatment or control groups to isolate the causal effect of a health policy.[39]	Estimates causal impact by comparing outcomes between randomly assigned groups.	Gold standard for internal validity; controls for both observed and unobserved confounders.	Often costly and time-consuming; ethical and logistical constraints in real-world health policy settings.
Interrupted Time Series (ITS)	Analyzes outcome trends before and after policy implementation to detect whether a significant shift occurs post-intervention, assuming prepolicy trends would have continued unchanged.[40]	Detects immediate and long-term changes in outcomes linked to policy timing.	Useful when randomization is not feasible; controls for baseline trends and seasonality.	Requires clear intervention timing; vulnerable to other simultaneous events ("history bias").
Difference-in-Differences (DiD)	Compares changes in outcomes over time between a policy-exposed group and a comparison group, assuming similar trends would have occurred without the policy.[41]	Isolates policy effects by subtracting out trends common to both treated and control groups.	Controls for time-invariant differences; useful with observational data and phased policies.	Requires strong parallel trends assumption; less reliable when groups differ in trend trajectories.
Regression Discontinuity Design (RDD)	Exploits a pre-determined eligibility cutoff (e.g., income, age) for assigning a policy to estimate impact by comparing those just above and below the threshold.[42]	Estimates local causal effects around the cutoff point.	High internal validity near the cutoff; avoids some ethical issues of randomization.	Results apply only to observations near the threshold; requires dense data around the cutoff.
Propensity Score Matching (PSM)	Matches individuals who received the policy with similar individuals who did not, based on observable characteristics, to estimate the treatment effect.[43]	Adjusts for confounding by balancing covariates across policy and comparison groups.	Controls for observable differences; intuitive to communicate and implement.	Does not address unobserved confounding; quality depends on matching quality and data richness.
Synthetic Control Method (SCM)	Constructs a weighted combination of untreated units to serve as a synthetic control group, against which the policy's effects on the treated unit are compared.[44]	Compares observed outcomes of treated unit to a counterfactual constructed from similar untreated units.	Useful for policy evaluation when no single control group is available; transparent and reproducible.	Requires extensive preintervention data; sensitive to selection of predictor variables.
Instrumental Variables	Uses an external factor (instrument) that influences policy exposure but is unrelated to the outcome, to identify causal effects.[45]	Estimates policy impact in situations where unmeasured confounding would bias other methods.	Enables causal inference without randomization when a valid instrument exists.	Finding a valid instrument is difficult; interpretation limited to compliers (local average treatment effect).

39 Zabor, E. C., Kaizer, A. M., and Hobbs, B. P. 2020. "Randomized controlled trials." *Chest* 158 (1 Suppl): S79–S87. https://doi.org/10.1016/j.chest.2020.03.013.
40 Lopez Bernal, J., Cummins, S., and Gasparrini, A. 2017. "Interrupted time series regression for the evaluation of public health interventions: A tutorial." *International Journal of Epidemiology* 46 (1): 348–355. https://doi.org/10.1093/ije/dyw098.
41 Wing, C., Simon, K., and Bello-Gomez, R. A. 2018. "Designing difference in difference studies: Best practices for public health policy research." *Annual Review of Public Health* 39: 453–469. https://doi.org/10.1146/annurev-publhealth-040617-013507.
42 Sasabuchi, Y. 2022. "Introduction to regression discontinuity design." *Annals of Clinical Epidemiology* 4 (1): 1–5. https://doi.org/10.37737/ace.22001.
43 Austin, P. C. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* 46 (3): 399–424. https://doi.org/10.1080/00273171.2011.568786.
44 Bonander, C., Humphreys, D., & Degli Esposti, M. (2021). Synthetic control methods for the evaluation of single-unit interventions in epidemiology: A tutorial. *American Journal of Epidemiology*, 190 (12), 2700–2711. https://doi.org/10.1093/aje/kwab211.
45 Iwashyna, T. J., & Kennedy, E. H. (2013). Instrumental variable analyses: Exploiting natural randomness to understand causal mechanisms. *Annals of the American Thoracic Society*, 10 (3), 255–260. https://doi.org/10.1513/AnnalsATS.201303-054FR.

Evaluation Method	Description	Analytical Focus	Strengths	Limitations
Timing of Events	Models the relationship between timing of policy exposure and timing of outcomes, capturing dynamic effects over time while accounting for observed and unobserved factors.	Examines temporal patterns in outcomes relative to policy engagement.	Captures anticipatory effects and delayed impacts; useful in staggered rollouts.	Computationally intensive; requires precise policy timing and long follow-up periods.

Source: Adapted from *Magenta Book*, page 48.

Value-for-Money evaluation methods[46]

In a health policy evaluation, several methods are commonly used to assess value-for-money by comparing the costs and benefits of different policies. These methods guide policy makers in determining the most efficient allocation of resources to achieve desired health outcomes.

To aid in selecting the most appropriate economic evaluation method, Table 10 provides a detailed overview of each method, including its key features, when to use it, pros, and cons.

Table 10 Overview of Commonly Used Value-for-Money Evaluation Methods

Evaluation Methods	Description	Analytical Focus	Strengths	Limitations
Cost-Effectiveness Analysis (CEA)	Compares the costs and health outcomes of one or more health policies. Estimates the cost to achieve a unit of health outcome—such as a life-year gained, QALY, or DALY—relative to another policy or the status quo,[47] for example, cost per life-year gained from a national cancer screening policy compared to no screening.	Measures the cost per unit of health outcome – for example, cost per life-year gained – enabling comparison of interventions targeting the same health outcome but differing in costs.	Useful for evaluating policies with common health outcomes; supports efficient resource allocation within the health sector; often based on clinical or epidemiological data; widely use in health technology assessments.	Cannot compare across policies with different outcome types, such as life-years gained vs. cases prevented; does not express health benefits in monetary terms, limiting use in cross-sector decisions – for example, life-years gained from tobacco control vs. maternal deaths averted through skilled birth attendance.
Cost-Benefit Analysis (CBA)	Evaluates whether the total benefits of a health policy, translated into monetary terms, exceed its costs, also expressed in monetary terms,[48] such as estimating the net economic return of a nationwide HPV vaccination policy, including reduced cancer treatment costs and productivity gains.	Examine the net economic value of a health policy by comparing total expected benefits and costs, allowing decision-makers to assess return on investment.	Facilitates comparison of policies across diverse sectors by standardizing both costs and benefits in monetary terms; provides a single monetary metric for decision-making; supports broad resource allocation decisions.	Monetizing health outcomes can be methodologically complex and ethically sensitive – for example, valuing the long-term benefits of preventing chronic diseases through early childhood nutrition policies.

Source: Adapted from *Magenta Book*, page 49.

Synthesis methods

Synthesis methods are vital in health policy evaluation as they allow for the integration of findings from various studies to form a comprehensive understanding of a policy's impact and its implementation. These methods help to address key evaluation questions by combining results from different methodologies, whether quantitative, qualitative, or mixed-methods. By synthesizing findings around specific evaluation questions, rather than merely presenting individual results, these methods aim to provide a more robust and cohesive evidence base—a

46 See the Economic Evaluation Types subsection for more detailed information.
47 World Bank. *Cost-effectiveness analysis*. DIME Wiki. https://dimewiki.worldbank.org/Cost-effectiveness_Analysis.
48 Independent Evaluation Group. 2010. *Cost-benefit analysis in World Bank projects*. World Bank. https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/cba_full_report1.pdf.

process often referred to as "triangulation." This approach can enhance the reliability of the evidence by providing a consensus where multiple sources align. However, if the evidence is conflicting, evaluators must carefully analyze discrepancies, consider alternative explanations, and seek additional data, if necessary.

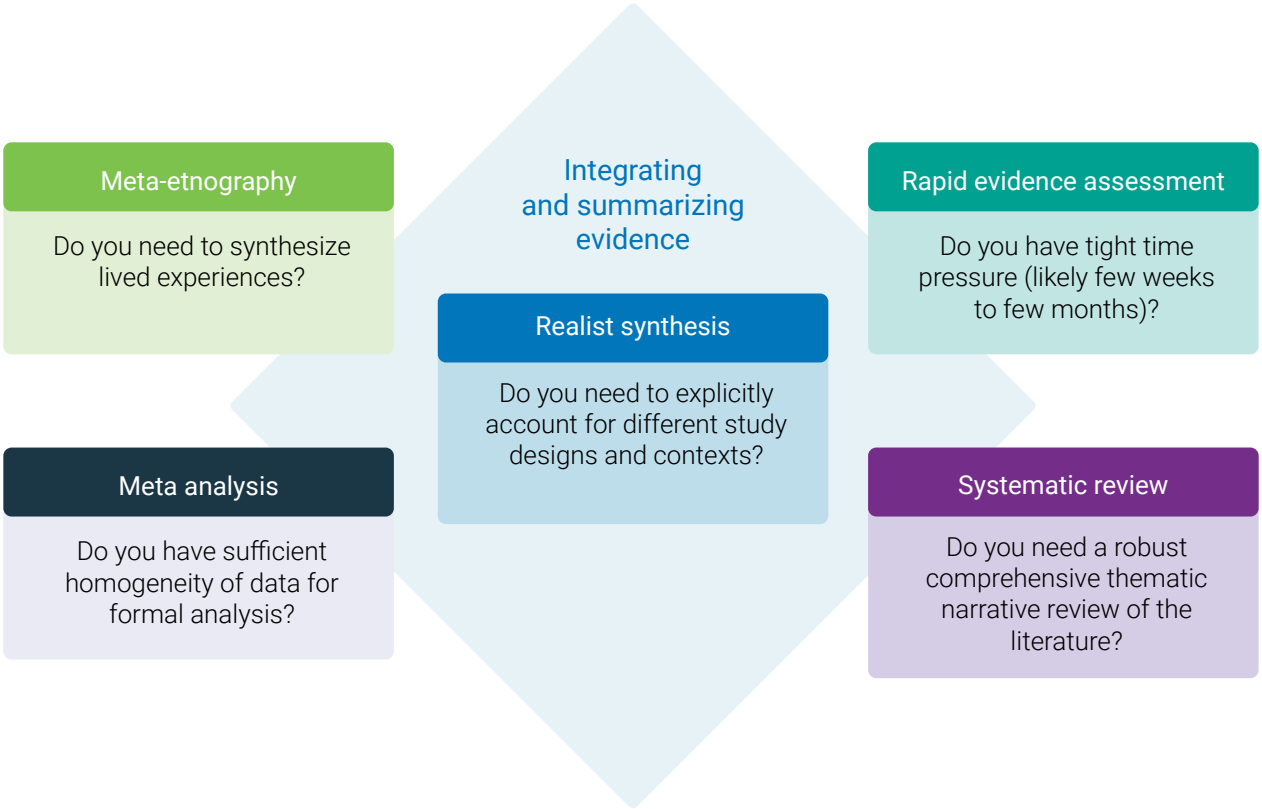
In addition to their role post-evaluation, synthesis methods can be employed at the pre-evaluation stage to integrate existing knowledge on a health policy topic. This helps to identify what is already known and highlights any gaps in the evidence base, guiding the design of new policies or interventions.

Formal synthesis techniques, such as meta-evaluation and meta-synthesis, are particularly useful to bring together the findings from multiple studies on the same policy issue. These methods involve systematically combining data to create a coherent narrative that addresses specific evaluation objectives. They rely on strict protocols and predefined criteria to assess the quality of included studies, ensuring that the synthesis is both comprehensive and credible.

A key challenge with synthesis methods is their reliance on existing data. The effectiveness of these methods is contingent on the availability and quality of prior studies, making rigorous inclusion criteria and thorough quality assessments essential. Systematic reviews and rapid evidence assessments are examples of synthesis methods that integrate existing literature to provide a narrative summary. While systematic reviews offer a thorough and methodical approach, they can be time-consuming, often requiring several months to complete. Rapid evidence assessments, on the other hand, provide quicker insights but may be less rigorous, making them suitable when decisions need to be made promptly—see Annex 3 for more information.

To assist evaluators in selecting the most appropriate synthesis method for their health policy evaluations, a set of guiding questions is provided below in Figure 7.

Figure 7 Selecting Evidence Synthesis Methods



Source: Adapted from *Magenta Book*, page 51.

After considering the guiding questions outlined in the figure above, evaluators should use Table 10 below to further refine their choice of synthesis method. The table provides a comprehensive overview of commonly used evidence synthesis methods, detailing their key characteristics, advantages, and potential limitations. This comparison will help evaluators select the most appropriate approach to effectively integrate and summarize evidence in their health policy assessments.

Table 11 Overview of Commonly Used Evidence Synthesis Methods

Evaluation Method	Description	Analytical Focus	Strengths	Limitations
Rapid Evidence Assessment (REA)	A streamlined literature review method that rapidly assesses the existing body of evidence on a focused policy question. Typically completed within 2–3 months using simplified protocols, and may optionally incorporate expert input.[49]	Rapidly scopes and synthesizes existing evidence to inform time-sensitive policy decisions.	Faster than traditional reviews; supports rapid decision-making; can identify evidence gaps for future research.	May omit relevant studies due to time constraints; less methodological depth and transparency than full systematic reviews.
Systematic Review	A rigorous method (mixed data) for identifying, appraising, and synthesizing all empirical evidence that meets prespecified criteria to answer a specific health policy question. It minimizes bias through structured and transparent protocols.[50]	Provides a comprehensive and unbiased synthesis of high-quality evidence to assess policy effectiveness.	High methodological rigor; minimizes bias; supports evidence-based policymaking.	Resource- and time-intensive (often ≥6 months); less feasible when evidence is sparse or decisions are urgent.
Meta-Analysis	A statistical technique (quantitative data) that combines results from multiple eligible studies to produce a pooled estimate of the effect of a health policy or intervention, improving precision and identifying patterns across data sets.[51]	Quantifies the magnitude of policy impact by statistically aggregating comparable findings from multiple studies.	Enhances precision; increases statistical power; reveals patterns or inconsistencies in study results.	Quality depends on original studies; requires methodological consistency; sensitive to publication bias and heterogeneity across studies.
Meta-Ethnography	A qualitative synthesis method that interprets and translates findings from multiple qualitative studies to generate new conceptual insights into health policy impacts, particularly from lived experiences.[52]	Interpretive synthesis of stakeholder perspectives to develop conceptual understanding of policy outcomes.	Generates context-rich, theory-informed insights; captures lived experience; useful for complex policy issues.	Dependent on quality of source studies; subjectivity in interpretation; limited generalizability.
Realist Synthesis	An interpretive review method that investigates how, why, and for whom health policies work (or do not) by analyzing context–mechanism–outcome (CMO) relationships. Combines realist evaluation principles with structured literature synthesis.[53]	Explores the causal mechanisms by which policies produce outcomes in specific contexts to build explanatory theory.	Unpacks complexity; explains how and why policies succeed or fail; adaptable to diverse implementation settings.	Requires domain and methodological expertise; time-consuming; not easily replicable or standardized.

Source: Adapted from *Magenta Book*, page 52.

Checklist for Step 3: Design Evaluation	
☐	Identify the evaluation methods, considering the data needs, level of comprehensiveness, and resource needs
☐	Utilize the appropriate method(s) to perform the policy evaluation

Step 4: Collect Data

Overview of Data Collection

Following the evaluation design, data collection becomes a critical component of the health policy evaluation process. It requires meticulous planning, as inadequate preparation or restricted access to data can render the

49 Crawford, C., Boyd, C., Jain, S., Khorsan, R., and Jonas, W. 2015. "Rapid Evidence Assessment of the Literature (REAL©): Streamlining the systematic review process and creating utility for evidence-based health care." *BMC Research Notes* 8: 631. https://doi.org/10.1186/s13104-015-1604-z.

50 Ahn, E., and Kang, H. 2018. "Introduction to systematic review and meta-analysis." *Korean Journal of Anesthesiology* 71 (2): 103–112. https://doi.org/10.4097/kjae.2018.71.2.103.

51 Ahn, E., and Kang, H. 2018. "Introduction to systematic review and meta-analysis."

52 Sattar, R., Lawton, R., Panagioti, M., and Johnson, J. 2021. "Meta-ethnography in healthcare research: A guide to using a meta-ethnographic approach for literature synthesis." *BMC Health Services Research* 21 (1): 50. https://doi.org/10.1186/s12913-020-06049-w.

53 Schick-Makaroff, K., MacDonald, M., Plummer, M., Burgess, J., and Neander, W. (2016). What synthesis methodology should I use? A review and analysis of approaches to research synthesis. *AIMS Public Health* 3 (1): 172–215. https://doi.org/10.3934/publichealth.2016.1.172.

evaluation infeasible, significantly constrained, or prohibitively costly. Furthermore, a poorly designed data collection strategy may result in the acquisition of inaccurate, incomplete, or irrelevant data, leading to flawed conclusions that compromise the validity and usefulness of the evaluation findings.

A key step in the data collection process is the gathering of baseline data—information collected before the policy is implemented. Baseline data provides a reference point for measuring the policy's impact over time. The data collection process begins with the identification of appropriate indicators, followed by determining data needs and sources, and ultimately gathering the necessary data. Effective data collection also requires managing logistics, conducting quality checks, and ensuring that data is collected in sufficient quantity and quality—see sample template in Annex 4.

Each of these elements contributes to collecting the right data, in the right amount, at the right quality, in the most efficient and effective way—topics that are further discussed in the following subsections.

How to Determine Indicators for Evaluation

The focus of the evaluation—identified through the policy description, stakeholder engagement, and the key questions developed during the evaluation design phase—is translated into measurable indicators during the data collection stage. An indicator is a specific, observable, and measurable characteristic or change that reflects progress toward achieving a policy's intended outcomes. Indicators are essential tools in the evaluation process: they determine what data will be collected, help to standardize assessment efforts, and ultimately inform evidence-based policy decisions. Indicators can be categorized based on their purpose—for example: process indicators, which track the implementation of activities; and outcome indicators, which measure the results or changes brought about by those activities.

Evaluators should either select existing indicators or develop new ones to address the evaluation focus and stakeholder questions. When feasible, using existing indicators is advantageous, as they are often validated, come with defined data sources, and support comparability across policies and contexts. Pretested indicators also improve the efficiency and credibility of the evaluation process. However, if existing indicators do not fully capture the specific processes, outcomes, or impacts of the policy in question, evaluators may develop new indicators, tailored to the policy's unique context and goals. Regardless of whether existing or new indicators are used, they should be aligned with or contribute to prioritized health indicators (Table 12). See also Annex 5.

To ensure quality and relevance, evaluators should follow established criteria for indicator selection or development. These criteria are detailed in Annex 6.

Table 12 Illustrative Indicators Developed from the Evaluation Findings of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners

List of indicators to be used for illustrative purposes		
Dimensions for assessment	**Policy questions to be answered**	**Indicators (illustrative) addressing the policy questions**
Patient safety	Does the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy contribute to improving patient safety?	% of patients' allegations against healthcare practitioners for negligence
		% of malpractice lawsuits per 1000 patients
		% of patients very highly or highly satisfied with patient safety in health services delivered by healthcare practitioners
		% of healthcare practitioners optimistic about the impact of the Malpractice Insurance Policy in reducing the number of medical errors
Quality of care	Does the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy contribute to improvement in healthcare quality?	Medical errors per 10,000 persons reported as a result of the negligence by healthcare practitioners
		% of clinical guidelines actively used in service delivery by healthcare practitioners
		% of healthcare practitioners using standard clinical guidelines in service delivery

List of indicators to be used for illustrative purposes		
Risk mitigation	Does the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy reduce financial burden for healthcare practitioners?	% of patients receiving malpractice compensation against the allegations for negligence of healthcare practitioners
		% of allegations patients received malpractice compensation corresponding to the severity of injury made by healthcare practitioner
		Amount of insurance payments per patient claiming for the malpractice compensation
	Does the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy reduce financial burden for healthcare practitioners?	% of healthcare professionals obtained malpractice insurance coverage
		% of healthcare providers sharing malpractice insurance premiums with healthcare practitioners
		% of annual malpractice insurance premiums in the total income of healthcare practitioners
		% of healthcare practitioners satisfied with the malpractice insurance benefits against costs
Profitability of MPI for insurance companies	Is the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy attractive for the insurance companies?	% of insurance payments in total malpractice insurance premiums
Access to healthcare services	Does the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy affect access to healthcare services?	% of healthcare practitioners withdrawing from the delivery of healthcare services with higher risks
Efficiency in health service delivery	Does the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy affect the efficiency of health service delivery?	Number of healthcare services per patient provided by healthcare practitioners

Source: Developed by the authors using evaluation findings from the implementation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (2024), in consultation with the Saudi Health Council.

What data sources should be considered, and how should they be collected for an evaluation?

Once the activities and outcomes to be measured have been identified and the indicators for tracking progress determined, selecting the appropriate data collection methods and sources becomes essential. A critical decision is whether existing data sources—secondary data collection—are sufficient to measure the indicators, or if new data must be gathered through primary data collection. Whenever possible, leveraging secondary data sources helps minimize both the financial costs of the evaluation and the burden on respondents. Common secondary data sources in health policy evaluation include:

- **Existing administrative and monitoring data**. Administrative data, such as healthcare resource information, services delivered, and beneficiary records, are often collected for operational purposes and can be repurposed for evaluation. Monitoring data, gathered throughout the policy cycle, help to track policy progress and can address policy questions related to inputs, processes, and outcomes.
- **National health surveys**. These surveys provide large, representative samples, offering valuable statistical insights into health behaviors and outcomes across populations.
- **Disease-specific health registries**. These registries contain detailed information about specific diseases, which can be useful for evaluating health policies targeting certain conditions.
- **Vital statistics**. Data from birth and death records, among others, provides critical information for evaluating the long-term impacts of health policies on population health.
- **Health surveillance systems**. These systems track disease outbreaks, health trends, and outcomes, making them crucial sources of information for health policy evaluations.
- **Hospital and clinic administrative data**. Patient data from healthcare facilities can be repurposed to assess health policy implementation and outcomes.
- **National population censuses and household surveys**. These broad datasets provide essential demographic and socio-economic information, relevant for evaluating the reach and equity of health policies.

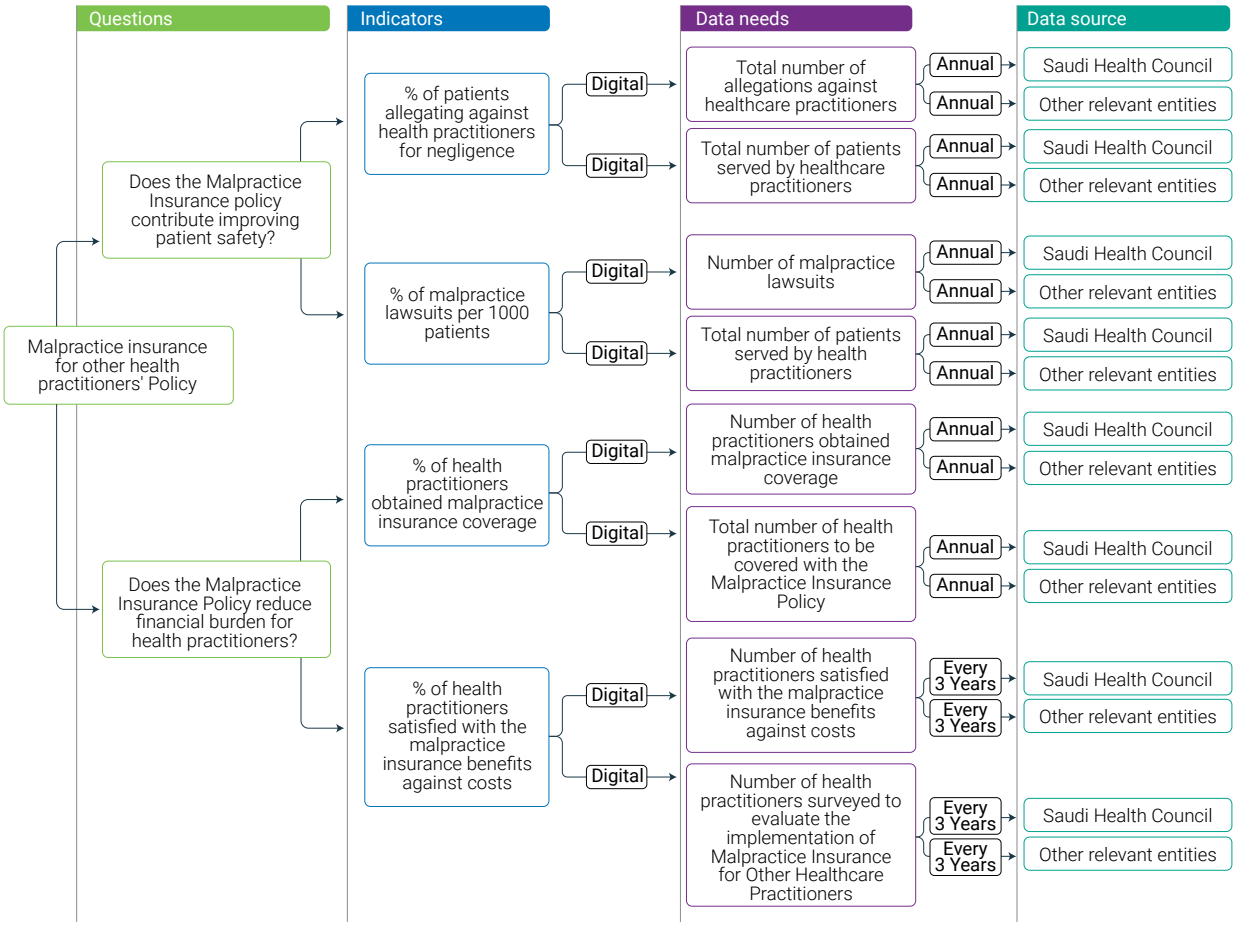
If secondary data sources do not adequately address the evaluation's needs, primary data collection methods may be necessary. Primary data collection methods also fall into several broad categories, such as:

- **New sources of data designed specifically for the evaluation**. These include tailored initiatives such as social media data collection to meet specific evaluation requirements.
- **New surveys conducted** for evaluation purposes, including personal interviews, telephone interviews, and instruments completed by respondents through mail or email.
- **Group discussions or focus groups**. These qualitative methods gather in-depth insights from specific populations or stakeholders.
- **Observation**. Observing behaviors and processes in real-time offers valuable data, especially for understanding how policies are being implemented.
- **Document reviews**. This involves reviewing records such as medical logs, diaries, or meeting minutes to gather relevant data for the evaluation.

Choosing the appropriate method from the available secondary and primary data collection options requires careful consideration of both context and content. Contextual factors include the budget available for data collection, the timeline for obtaining results, and any ethical considerations. Content-related questions must also be addressed, such as whether the topic involves sensitive issues, observable behaviors, or information the respondent is likely to know.

While the primary focus is on selecting the most relevant data sources and methods, mapping out how the evaluation questions, indicators, and data collection methods relate to each other can offer additional clarity—see Figure 8. This type of visualization provides an optional tool to help evaluators to organize the relationships between various data sources and formats, allowing for easier tracking and understanding of how each part of the evaluation fits together.

Figure 8 Illustrative Data Flow and Mapping Visualization Developed from the Evaluation Findings of the Mandatory Medical Malpractice Insurance Policy for Health Practitioners



Source: Developed by the authors using evaluation findings from the implementation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (2024), in consultation with the Saudi Health Council.

Employing Multiple and Mixed Methods in Data Collection

In many cases, a single data collection method may not adequately capture the complexity of a health policy, or the full range of outcomes being assessed. In such instances, employing multiple or mixed methods of data collection can enhance the depth and credibility of the evaluation.

Using mixed methods—the deliberate integration of both qualitative and quantitative approaches—allows evaluators to explore not only what changes occurred, but also how and why they happened. This approach can improve the validity of findings and provide a more comprehensive understanding of the policy's effects. One common benefit of mixed methods is triangulation, where findings from different methods are cross validated to increase the reliability and robustness of the conclusions.

There are two typical designs for using mixed methods in evaluation:

1. **Sequential mixed methods**. One method is used first to inform or complement the next. For example, focus **groups** (qualitative) may be conducted to inform the design of a survey instrument (quantitative), followed by follow-up interviews (qualitative or mixed) to explore findings in greater depth.
2. **Concurrent mixed methods**. Both qualitative and quantitative methods are applied in parallel. For example, focus groups or open-ended interviews may be conducted alongside a quantitative survey to confirm or enrich the interpretation of survey responses.

By employing mixed methods, evaluators can address complex evaluation questions more effectively, enhance stakeholder confidence, and strengthen the overall rigor of the health policy evaluation. This approach is especially valuable when the outcomes being evaluated are abstract or when no single high-quality data source exists, as it maximizes the strengths and mitigates the limitations of each method.[54]

How to Ensure Data Quality

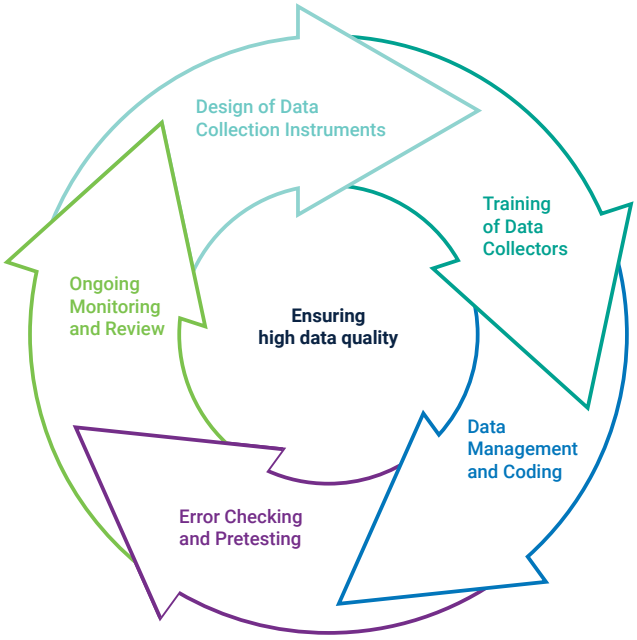
Ensuring data quality is essential to produce valid, reliable, and actionable evaluation findings. High-quality data allows stakeholders to trust the conclusions of the evaluation and use the results for informed decision-making. In a health policy evaluation, data quality encompasses several key factors: *reliability, validity, accuracy, consistency, and completeness*. Each of these components must be carefully addressed to maintain the credibility of the evaluation process and its findings.

- **Reliability**: This refers to the consistency of data over time and across different data collectors or sources. High reliability ensures that repeated measurements under the same conditions produce the same results. For example, if different evaluators collect data from the same health policy intervention, the outcomes should remain consistent across time, methods, or personnel.
- **Validity**: This determines whether the data accurately reflects what it is intended to measure. This factor is crucial in health policy evaluation as it ensures that data supports valid conclusions about the policy's effects on health outcomes.
- **Accuracy**: Data must be precise and error-free. Accurate data accurately reflects real-world conditions and policy impacts, avoiding bias or misrepresentation.
- **Consistency**: This refers to ensuring that data remains stable across different contexts and time frames. For example, data collected in one phase of a health policy should align with data collected in other phases to allow for meaningful comparisons over time.
- **Completeness**: Complete data ensures that no critical information is missing, which can impact the evaluation's conclusions. Incomplete data can lead to biased results and misinterpretation of the policy's effects.

To achieve high-quality data, evaluators must follow a series of steps designed to enhance the factors influencing data quality. Figure 9 shows the essential steps that ensure that these factors are realized during data collection and analysis.

54 *CDC's Program Evaluation Framework Action Guide*, page 63.

Figure 9 Steps in Ensuring Data Quality



Source: Authors.

1. Design data collection instruments	Design well-structured data collection instruments that align with the evaluation's goals. Questions should be clearly worded to avoid ambiguity, and the instruments should be pretested to identify any issues.
2. Train data collectors	Proper training is vital to ensure that data collectors understand the protocols and methods for gathering data consistently. Training should cover both technical aspects (such as using data collection tools) and ethical considerations (such as confidentiality).
3. Data management and coding	Proper data management practices should be implemented to ensure that data is recorded, stored, and retrieved in an organized manner. This includes coding data systematically to allow for easy analysis and retrieval.
4. Error checking and pretesting	Conduct error checks throughout the data collection and entry process to identify inconsistencies or mistakes. Pretesting data collection tools allows evaluators to identify potential issues before large-scale data collection begins, improving data accuracy.
5. Ongoing monitoring and review	Regularly review the data quality throughout the evaluation process. Continuous monitoring can help to identify and resolve any data quality issues before they affect the final analysis.

By following these steps, evaluators can ensure that the data used in health policy evaluations is of high quality—reliable, valid, accurate, consistent, and complete—which, in turn, contributes to more credible, transparent, and actionable conclusions.

What Ethical Considerations Need to Be Addressed in Data Collection and Quality Assurance?

It is important to address the ethical considerations involved in data collection and analysis. Ethical data practices help to maintain trust with stakeholders and ensure that the evaluation process respects privacy, confidentiality, and integrity. Following are key considerations.

1. Confidentiality	Protecting the privacy of individuals whose data are collected is paramount. Ensure that all personal and sensitive information is kept confidential, and only authorized personnel have access to it. Data anonymization techniques should be used where applicable.
2. Informed consent	Participants in health policy evaluations must provide informed consent, understanding the purpose of data collection, how their data will be used, and their right to withdraw from the study at any time without penalty. Consent forms should be clear, accessible, and culturally appropriate—see Annex 7 for a sample.

3. Data security Strong data security measures should be implemented to protect against unauthorized access or data breaches. This includes encrypting data, restricting access, and ensuring that data are stored in secure systems. Evaluators should also be prepared to handle any potential ethical dilemmas that arise during data collection, such as breaches of confidentiality or unanticipated harms to participants.

By consistently applying these standards and following these steps, evaluators deepen their understanding of a health policy context and significantly enhance the effectiveness and impact of their evaluations.

How Much Data Should Be Collected?

Determining the appropriate quantity of data is essential to ensuring accurate and reliable conclusions. While some evaluations require data of the highest validity and reliability, particularly when supplemented by research studies, there are instances where a smaller sample or convenience sampling may suffice. The following factors should guide the decision on the quantity of data to collect.

1. **Evaluation questions:** The data requirements depend on the evaluation's scope. For example, evaluations focusing solely on policy-related questions may require fewer data points than those with broader research objectives. Additionally, quantitative evaluations generally require more data than qualitative evaluations. The sample size will vary based on the level of detail needed and the types of comparisons being made – for example, comparing different population groups or time periods.
2. **Level of jurisdiction:** The scale of the evaluation—national, regional, or local—significantly impacts the amount of data needed. Evaluations at higher jurisdictional levels, such as national or regional, typically require larger and more representative samples due to greater population heterogeneity. Probability sampling methods, such as stratified or multistage sampling, may be necessary to ensure that the data is representative of the broader population. These methods help to account for population diversity and improve the reliability of the findings. In contrast, local or community-level evaluations generally involve smaller samples, which are less costly and can still provide sufficient insights. In these cases, non-probability sampling methods like convenience or purposive sampling may be more appropriate when generalizability is less of a concern, but the limitations of such methods should be noted.
3. **Size of change (effect size):** Evaluations aiming to detect small changes in health outcomes or policy impacts typically require larger sample sizes. This is because small changes are harder to detect and require more data to ensure statistical power. Effect size refers to the anticipated magnitude of the policy's impact, such as a 5% or 10% reduction in hospital readmissions. The smaller the expected effect size, the larger the sample size needed to detect it with confidence.
4. **Statistical power and power calculations:** To ensure that the sample size is adequate, power calculations should be used. These calculations help evaluators to determine how much data is needed to achieve a statistically sound evaluation. Key elements of power calculations include
   - » **Effect size:** The anticipated impact of the policy – for example, reduction in hospital readmissions by 5%.
   - » **Significance level:** Usually set at 0.05, this represents the likelihood of incorrectly concluding that the policy had an effect when it did not (Type I error).
   - » **Desired power level:** Typically set at 80%, meaning there is an 80 percent chance of detecting a true effect if it exists, thus minimizing the risk of missing significant outcomes (Type II error).

   Power calculations balance data sufficiency with resource efficiency. For example, detecting smaller changes in health outcomes requires larger sample sizes, while more significant changes can be identified with fewer data. By conducting power analysis, evaluators can ensure that they collect enough data to avoid Type I errors (false positives) and Type II errors (false negatives).[55]
5. **Adjusting for data loss:** Evaluators should anticipate potential data loss, such as incomplete responses or dropouts, and adjust the sample size accordingly. This is often done by inflating the calculated sample size to maintain the study's statistical validity. Alternatively, methods like data imputation can be used to address missing data, ensuring that conclusions remain reliable despite incomplete datasets.

55 In the context of health policy evaluation, it's important to recognize the potential for **Type I** and **Type II errors** during the analysis phase:
- Type I Error (False Positive): This occurs when an evaluation wrongly concludes that a health policy had an effect when it did not. For example, an evaluation might mistakenly attribute a reduction in hospital readmissions to a newly implemented policy, when in fact the change was due to unrelated external factors. This can lead to the continuation of ineffective policies based on incorrect conclusions.
- Type II Error (False Negative): This happens when an evaluation fails to detect a true effect of the policy. For instance, if a public health campaign significantly improved vaccination rates, but the evaluation did not have sufficient data or statistical power to detect the change, the policy's effectiveness may be overlooked. This can result in the abandonment or underutilization of effective policies.

Evaluators must be mindful of these errors and use power calculations to ensure that they collect enough data to make accurate, well-supported conclusions about the policy's true impact.

How to handle data collection

Appropriate data collection and handling are fundamental to ensuring the integrity and security of health policy evaluations. All data used in an evaluation, regardless of its source, must be collected, transferred, stored, processed, and deleted in accordance with national laws and relevant security processes.

Data Security and Access Protocols

It is critical that everyone handling data is properly trained to understand data security procedures. Data access protocols should clearly define who has authority to access the data, how remote access is managed, and the specific requirements for handling and securing sensitive information. Some key considerations include:

- Authorizing data access for different user groups based on roles and responsibilities.
- Implementing security measures for remote access, including the use of secure Wi-Fi.
- Conducting mandatory data handling and security training.
- Masking or encrypting personal data to ensure privacy.

Data should be stored in secure formats, such as databases, spreadsheets, or data warehouses. Access to this data should be limited to those with legal entitlement and a work-related need, ensuring that only authorized personnel can read, write, or modify the information.

Anonymization and De-identification of Data

To safeguard privacy, data should be anonymized as early as possible during the evaluation process. This involves removing direct identifiers, such as names or addresses, from the analytical dataset. If retaining this information is temporarily necessary, it should be stored separately with strictly controlled access. Even with anonymization, data may still qualify as personal data, necessitating careful handling under data protection laws.

Evaluators should ensure full alignment with national efforts to promote cybersecurity and ethical data governance. In the Kingdom of Saudi Arabia, this includes adhering to the Personal Data Protection Law and relevant national bioethics frameworks. Key references include the National Committee of Bioethics (NCBE)[56] and the Personal Data Protection Law.[57] These frameworks guide the ethical handling of sensitive data and uphold the principles of privacy, confidentiality, and responsible data use in research and evaluation context.

At the conclusion of an evaluation, every effort should be made to fully anonymize the data. This could involve replacing precise identifiers such as date of birth with broader categories like year of birth, or substituting exact locations with general geographic regions. If full anonymization is not feasible and personal data must be retained, appropriate safeguards must be established, including secure storage and restricted access.

Risks of Improper Data Handling

The risks associated with improper data handling are significant and include breaches of confidentiality or data security, harm to individuals or communities through privacy violations, and legal challenges or reputational damage to the department or government. Mitigating these risks requires adhering to strict data handling protocols and regularly reviewing compliance with legal and organizational standards.

Agreements and Roles in Data Collection

Agreements among the parties involved in the evaluation process are essential for clarifying roles, responsibilities, and procedures for managing data and conducting the evaluation. These agreements may take the form of legal contracts, memoranda of understanding, or detailed protocols. Elements of an agreement include:

- Intended users and uses of the data.
- The purpose of the evaluation and key evaluation questions.
- Evaluation design and methods.
- A summary of the deliverables, timeline, and budget.

Creating these agreements helps establish a shared understanding of the evaluation activities and data handling procedures. It also provides a reference point for any necessary modifications during the evaluation process.

56 National Committee of Bioethics (NCBE). "About the Committee." King Abdulaziz City for Science and Technology (KACST). Retrieved from https://ncbe.kacst.edu.sa.

57 Saudi Data and Artificial Intelligence Authority (SDAIA). 2022. *Personal Data Protection Law*. https://sdaia.gov.sa/ar/SDAIA/about/Files/RegulationsAndPolicies02.pdf.

Agreements ensure that all parties involved are aligned on the data security measures, access protocols, and other logistical aspects of the evaluation.

Checklist for Step 4: Data collection and analysis	
☐	Identify the indicators required to measure the identified evaluation focus and evaluation questions
☐	Decide to choose the indicators from the existing inventory or to develop new ones.
☐	Consider the availability of data and choose the most appropriate data source(s).
☐	Review various data collection methods and select those that are best suited to the context and content.
☐	Pilot test new newly developed indicators to identify and control potential sources of error.
☐	Consider adopting a mixed-method approach for data collection.
☐	Evaluate both quality and quantity issues related to the data collection process and adjust to the available resources and time.
☐	Create a detailed protocol to guide the data collection process.

Step 5: Justify Conclusions

The next step in health policy evaluation is the justification of conclusions, which increases the likelihood to use the evaluation results for decision-making. At this stage, policy evaluators analyze the data, make assertions regarding a policy's outcomes and impacts, and justify these conclusions. Justifying conclusions involves analyzing and synthesizing data, establishing performance standards for the health policy, interpreting the results, and making well-supported claims that align with the expectations of the health system and its stakeholders. When stakeholders—who may have differing perspectives on what constitutes success of a policy—see if the conclusions reflect their values and the policy's intended goals, they are more inclined to use the findings to guide decisions.[58]

At this stage the indicators are reviewed, results are tabulated, outcomes are compared and presented in a way that is easily understandable by policy makers, stakeholders, and evaluators. The goal is to present a clear narrative of the policy's effectiveness and impact, based on the evaluation findings.

The process of analysis and justification typically goes through the following key stages:

1. **Data entry and error checking.** Enter the collected data into a secure system or database, ensuring that it is accurate and free of errors. Health policy evaluations often utilize existing systems like health information systems or administrative datasets, but if new data are collected – for example, surveys – it is essential to use appropriate software to enter, check, and accurately tabulate the data.
2. **Tabulate data.** For each indicator, calculate basic statistics to generate key insights, such as the total number of participants or health beneficiaries, the number of participants who experienced the intended health outcome, and the percentage of participants or regions achieving the policy's desired outcomes.
3. **Stratify data by key variables.** Analyze the data by breaking down into relevant demographic or geographic categories, such as age, gender, socioeconomic status, or region. This stratification allows evaluators to assess whether the policy had different impacts across subgroups, ensuring that the policy's outcomes are equitable.
4. **Make comparisons.** In health policy evaluation, it is important to compare data across different groups, regions, or time periods. This could involve comparing intervention groups with control groups, pre-policy vs. post-policy outcomes, or regional variations in policy impact. Statistical tests may be used to show differences between these groups, helping to identify where the policy was most or least effective.
5. **Present data in an understandable format.** Use clear visual tools, such as bar charts, pie charts, line graphs, and maps to present the evaluation data. Policy makers and stakeholders need to easily interpret the findings to make informed decisions about the health policy's future. Visual data presentations help to highlight key trends and outcomes, making complex data more accessible.

58 *CDC's Program Evaluation Framework Action Guide*, page 74.

Sample Case Survey to Assess the Implementation of the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy

Survey overview

Saudi Health Council conducted a survey, which aimed to provide a comprehensive assessment of this policy among other health practitioners by measuring the compliance of healthcare providing entities, the number of insured other healthcare workers, and the level of awareness and understanding of the policy across the health practitioners.

Data entry and error checking

The evaluation team collected the survey data based on the standardized questionnaires which were designed for health practitioners and healthcare providing facilities separately. To ensure the accuracy and the objectivity of the results the evaluation team used IBM® SPSS statistical software in analyzing data and employed advanced statistical methods.

Tabulate data

A total of 261 responses were collected from the healthcare facilities, which included a variety of institutions such as hospitals, clinics, primary care centers, polyclinics, and optical shops. Additionally, 1,112 healthcare practitioners responded to the questionnaire. After excluding physicians and dental assistants, the sample consisted of 1,090 healthcare practitioners. Following the removal of duplicate responses, the final sample size was 1,069 healthcare practitioners. The survey provides basic statistics including the frequencies and percentages across the participants.

Stratify data by key variables

The health practitioners participating in the survey were demographically stratified to gender and age groups to assess the policy's impact across different demographic groups. Out of 1,069 participants, 556 were males, making up 52.0% of the sample, while the number of females was 513, representing 48.0%.

Among the 1,069 participants, the 35–44 years age group constituted the largest group, with 437 participants, equivalent to 40.9% of the total. This was followed by the 25–34 years age group with 419 participants, representing 39.2%. The youngest age group (18–24 years) had 34 participants, accounting for 3.2%. The 55–64 years age group comprised 35 participants, making up 3.3%. There were only two participants aged 65 years or older, 0.2% of the total.

The participants were also stratified to geographic locations across 13 regions. Riyadh region had the largest number, 304 participants, representing 28.4% of the total. This was followed by Makkah region with 201 participants, accounting for 18.8%, and the Eastern region with 183 participants, making up 17.1%. Al-Baha region had only 12 participants, representing 1.1%, making it the region with the lowest percentage among the regions of the Kingdom.

Additionally, the survey designers broke down the participants into groups according to the level of their education. Thus, the data showed that those with a bachelor's degree represented the largest proportion at 57.25%. They were followed by those with a diploma, at 32.27%. Others were: participants with a master's degree, 6.74%; a doctoral degree, 2.06%; and a board certification, 1.68%.

Make comparisons

Various data analysis methods were used to understand the general distribution of the data, including descriptive analysis that covered basic statistics such as means, frequencies, and percentages. Inferential analysis was also conducted through hypothesis tests such as the T-test and chi-square test to determine if statistically significant differences existed between groups.

Present data in an understandable format

The survey results were presented in formats such as tables, bar charts, and pie charts to compare the results across groups identified. The different formats helped to cross-check various data to produce indicators.

Source: Developed by the authors using evaluation findings from the implementation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (2024), in consultation with the Saudi Health Council.

In evaluations that use multiple methods (quantitative and qualitative), it is important to combine various sources of evidence to reach a comprehensive understanding of the policy's impact. This involves detecting patterns in the data, integrating qualitative insights, such as from focus groups or interviews, and comparing these with the quantitative findings like health outcomes or cost savings. For example, if a policy aimed at improving access to healthcare shows a quantitative improvement in service utilization, qualitative data from healthcare providers might provide insight into barriers that remain or unintended effects of the policy. This synthesis helps evaluators not only to assess whether the policy met its objectives but also to understand why certain outcomes occurred, providing actionable recommendations for improvement.

In the evaluation process, performance benchmarks[59] are also essential tools used to assess the effectiveness and impact of policies. These benchmarks reflect the values and expectations of stakeholders regarding the health policy and are fundamental to sound evaluation practices. Policy makers and stakeholders must work together to define the criteria that will be used to determine whether a policy is deemed successful, adequate, or in need of improvement. This process ensures that the evaluation aligns with national health priorities and the needs of the population.

When conflicting views arise about the quality, value, or significance of a policy, it often suggests that stakeholders are employing different standards or values. Such disagreements should encourage stakeholders to clarify their values and work toward a consensus on how to interpret the evaluation findings. To prevent such conflicts, it is critical to establish clear performance benchmarks for each indicator early in the evaluation process—see Step 1: Identify and Engage Stakeholders. These benchmarks are often based on expected improvements from baseline data and should be realistic yet aspirational, considering the policy's maturity, the capacity of the health system, and the priorities of the stakeholders involved. Early agreement on benchmarks helps to prevent disputes later in the process and contributes to a smoother, more effective evaluation.

Evaluators can apply the following best practices to accurately interpret and communicate evaluation findings:

- Interpret results in relation to the policy's original goals and objectives.
- Tailor the reporting to the intended audience, ensuring that findings are relevant and understandable to policy makers, implementers, and the public.
- Acknowledge limitations, including
  - » Potential sources of bias
  - » Limitations in data validity or reliability
  - » Constraints in study design or context
- Consider alternative explanations for the observed results.
- Compare findings with those relating to similar policies or interventions in other contexts to identify trends or anomalies.
- Triangulate results across data collection methods to assess consistency and credibility.
- Assess alignment with established theories and previous research findings.
- Check whether the results align with expectations, based on prior assumptions, baseline data, or logic models of the policy.

The set of questions in Table 13 offers a practical tool to enhance the acceptability of judgments made based on the findings of the policy evaluation.

Table 13 Acceptability Check of Evaluation Findings

#	Questions	Response
1	Who will analyze the data (and who will coordinate this effort)?	
2	How will data be analyzed and displayed?	
3	Against what values and benchmarks will the interpretations be compared in forming the judgments?	
4	Who will be involved in making interpretations and judgments, and what process will be employed?	
5	How will the conflicting interpretations and judgments be dealt with?	

59 A performance benchmark in health policy evaluation is a specific, measurable standard or target used to assess the effectiveness or efficiency of a health policy. It serves as a reference point to determine whether the policy has achieved its intended outcomes or to compare its performance against established norms or peers, often based on historical data, best practices, or policy goals.

6	Are the results like what was expected? If not, why are they different?
7	Are there alternative explanations for the results?
8	How do your results compare with those of similar programs?
9	What are the limitations of the data analysis and interpretation process – for example, potential biases, generalizability of results, reliability, validity)?
10	If multiple indicators were used to answer the same evaluation question, were the results received similar?
11	How will the findings be communicated to ensure stakeholders interpret them accurately and effectively for decision-making?

Source: CDC's Program Evaluation Framework Action Guide, page 80.

Checklist for Step 5: Checklist - Justify Conclusions	
☐	Consider context when analyzing data and make judgments.
☐	Assess results against available literature and those of similar policies.
☐	Consider alternative explanations of the findings.
☐	Use existing values and benchmarks – for example, KSA Vision for 2030 and Health Sector Transformation Program – as a starting point for comparisons.
☐	Compare the health policy outcomes with those of previous years.
☐	Compare actual findings with intended outcomes of the health policy.
☐	Document potential biases.
☐	Examine the limitations of the evaluation.

Step 6: Use and Disseminate Evaluation Findings

What is the value of using and disseminating the evaluation findings?

The true value of a health policy evaluation is realized through its use in decision-making. Since the primary goal of any policy evaluation is to apply the insights gained in enhancing or reforming health policies, this should be a priority from the planning stage of the evaluation and continuously revisited throughout the process. The purpose(s) identified early on must guide how the findings are applied to support evidence-based health policy decisions.

Evaluations act as critical feedback mechanisms, helping to guide the development of new policies, refine existing ones, and justify resource allocation. Effective use of health policy evaluations means embedding insights and conclusions into the broader policy-making process, fostering accountability, and creating incentives for continuous improvement. If the results are not used, there is a disconnect between the evidence gathered and the real-world policy decisions, resulting in missed opportunities for learning as well as applying lessons to facilitate improved health outcomes.

Evaluation findings are used more specifically in strengthening health policies by providing evidence of their effectiveness and progress toward health goals. They also help to identify areas for improvement in policy design and implementation, ensuring that strategies remain adaptive and responsive to the evolving needs of the health sector. Moreover, evaluation results are valuable tools to justify existing funding and advocate for increased investments in health programs. They assist in informed budget planning and allocation, highlight urgent health priorities, and enhance communication with stakeholders, including the public, healthcare providers, and policy makers.

Users of health policy evaluation findings can be categorized into direct and indirect users.

- **Direct users:** These are the stakeholders directly involved in the design, implementation, and oversight of the health policy. They use the findings to refine policies, improve their implementation, and ensure that intended health outcomes are achieved. For example, health sector policy makers, public health agencies, and program managers directly apply evaluation results to make informed decisions.
- **Indirect users:** These include the external stakeholders who use the evaluation findings to inform future health policy design, improve similar programs, or ensure the responsible use of public funds. For instance, international organizations, healthcare funders, researchers, or advocacy groups may apply the lessons learned to new policies or campaigns in similar health domains.

How to maximize the use of evaluation findings

To maximize the use of the evaluation findings, the following five essential elements should be considered:[60]

1. **Recommendations.** These are suggested actions derived from the evaluation findings. They can enhance the evaluation's impact if they are aligned with stakeholders' concerns and supported by strong evidence. However, if recommendations lack sufficient backing or do not resonate with stakeholder values, they can undermine the evaluation's credibility. The relevance and usefulness of recommendations vary depending on the target audience and the evaluation's goals. Actions taken during earlier stages, such as stakeholder identification, policy design, and data collection, help to ensure that the recommendations are pertinent and beneficial to all parties involved.
2. **Preparation.** This involves the steps taken to facilitate the practical use of evaluation findings. Through adequate preparation, stakeholders can:
    - » Improve their ability to apply new insights effectively.
    - » Consider the potential influence of the findings on decision-making processes.
    - » Analyze the possible positive and negative impacts of the findings and identify options for policy enhancement.
3. **Feedback.** This is the continuous exchange of information among all participants in the evaluation process. It is essential at all stages to build trust and maintain alignment with the evaluation's objectives. Early feedback helps keep the evaluation on track by keeping everyone informed about the policy implementation and the evaluation's progress. As preliminary findings become available, feedback allows stakeholders to provide input on important decisions. Effective feedback is typically gathered through discussions, sharing interim results, and reviewing draft reports.
4. **Follow-up.** It is a means to provide ongoing support to users after they receive the evaluation findings and start drawing conclusions. Active follow-up can remind users of the intended use of the findings, ensure that the results are applied to the core evaluation questions, prevent misapplication, and safeguard against the loss or neglect of important lessons during complex or politically charged decision-making.
5. **Dissemination.** This refers to the process of sharing the evaluation findings and methodologies with relevant audiences in a timely and impartial manner. The aim is to achieve full transparency and unbiased reporting. Effective dissemination requires early discussions with intended users and stakeholders about the reporting strategy and adapting the timing, style, tone, source, medium, and format of the information to suit the needs of the audience. Identifying and using the most appropriate communication channels and formats is essential to ensure that the evaluation findings reach and effectively impact all stakeholder groups.

Sample Case **Recommendations Based on the Study Results to Assess the Implementation of the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy**

Recommendations Based on Study Results

The mandatory insurance policy against medical errors is an important step aimed at enhancing patient safety and protecting healthcare practitioners. Based on the results to date, the following recommendations have been formulated to ensure the effective implementation of this policy.

- Mandate healthcare service providers to verify the insurance coverage of healthcare practitioners.
- Introduce supplemental policies to support implementation—such as subsidies for insurance premiums, a phased implementation timeline, and other financial relief mechanisms for specific practitioner categories.
- Review insurance prices in line with the specializations of healthcare practitioners, to avoid imposing additional financial burdens.
- Collaborate with insurance companies to review and standardize the content of malpractice insurance policies, ensuring clarity, consistency, and alignment with policy requirements.
- Study the possibility of exempting some healthcare practitioners, such as those not working in the healthcare providers, from mandatory insurance requirements to ensure no unnecessary burdens are imposed.

60 *CDC's Program Evaluation Framework Action Guide*, page 83.

- Organize workshops and training courses to increase practitioners' awareness of the importance of insurance and its benefits, as well as clarify the related policies and procedures.
- Regulate the establishment of new insurance companies focusing on the policy to increase access and to foster fair competition.

Implementing these recommendations in a scientific and studied manner will enhance the effectiveness of the mandatory insurance policy against medical errors, contributing to better healthcare and protecting the rights of practitioners and patients alike.

Source: Developed by the authors using evaluation findings from the implementation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (2024), in consultation with the Saudi Health Council.

How to disseminate evaluation findings

To ensure that evaluation findings are effectively used, it is essential to communicate these findings to stakeholders and make them publicly accessible.[61] Evaluations that clearly articulate evidence-based implications for future decision-making only hold value if decision-makers can access and use the findings promptly. Similarly, oversight bodies and the public should be able to easily access, comprehend, and apply these findings to evaluate government policy design, implementation, and outcomes.

Disseminating evaluation findings can be done through various channels and formats. The presentation of evidence should be strategic, guided by the evaluation's objectives, and tailored to the information needs of the intended users.[62] When evaluation results are well-synthesized, customized for specific audiences, and directly delivered to them, it enhances their usability. Tailored communication and dissemination strategies are crucial for ensuring that stakeholders have easy access to the findings, which greatly increases the likelihood that the findings will be used.

One common method is to prepare and distribute an evaluation report. The report should clearly and impartially convey all aspects of the evaluation, presented in a concise and understandable manner. The report does not need to be overly technical or lengthy. A typical structure for such a report might include:[63]

- **Executive Summary**
- **Background and Purpose**
  - » Policy background
  - » Evaluation rationale
  - » Stakeholder identification and engagement
  - » Policy description
  - » Key evaluation questions/focus
- **Evaluation Methods**
  - » Design
  - » Sampling procedures
  - » Measures or indicators
  - » Data collection procedures
  - » Data processing procedures
  - » Analysis
  - » Limitations
- **Results**
- **Discussion and Recommendations**

Effective communication of evaluation findings can also involve diverse methods such as one-page summaries, videos, infographics, data dissemination, newsletters, social media updates, and presentations at conferences and seminars. Engaging with stakeholders to understand their specific evidence needs can guide the selection of the most appropriate communication channels. While it is important to not discuss findings publicly before they have been formally published, involving key stakeholders early, especially with negative findings, can help

61 OECD 2020, page 128.
62 OECD 2020, page 130.
63 *CDC's Program Evaluation Framework Action Guide*, page 86.

to manage the messaging effectively before public dissemination.

A comprehensive dissemination plan for evaluation findings should be developed and implemented to coordinate these efforts more effectively. Such a plan would detail:[64]

- The list of stakeholders and evaluation users.
- Specific information to be shared with identified stakeholders.
- The timing and purpose of the information dissemination, customized to the needs of different users.
- The format—business, scientific, briefs, single issue/topic or multiple; choice of communication tools—print, online, social media; schedule for publishing information—quarterly, annually, or during seminars and conferences.
- Coverage of all outputs to be published, including reports, underlying data, and research methodologies.

The plan, as outlined in Table 14 should also encompass strategies for enhancing accessibility and engagement, such as using infographics, creating executive summaries, distributing "information nuggets" through social media, and organizing seminars to present findings. This approach ensures that all relevant stakeholders have timely and appropriate access to the evaluation insights, facilitating informed decision-making and policy refinement.

Table 14 Dissemination Plan (Sample case from the Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy)[65]

Stakeholders / Users	What information do they want?	How do they want this information (format/ communication tool)?	When do they want this information?	Who is responsible for sharing this information?
Ministry of Health	Does the policy contribute to improving patient safety?	Report	Annual	SHC
	Does the policy contribute to healthcare quality improvement?	Report	Annual	SHC
	Does the policy affect access to healthcare services?	Report	Annual	SHC
	Does the policy affect the efficiency of health service delivery?	Report	Annual	SHC
Patient Safety Center	Does the policy contribute to improving patient safety?	Report	Annual	SHC
	Does the policy contribute to healthcare quality improvement?	Report	Annual	SHC
Saudi Commission for Health Specialties	Does the policy reduce the financial burden for healthcare practitioners?	Report	Annual	SHC
	Does the policy reduce financial burden for healthcare providing facilities?	Report	Annual	SHC
Patients	Does the policy contribute to improving patient safety?	Press release / social media / policy briefs etc.	Regular	SHC
Private Sector	Is the policy attractive for insurance companies?	Report	Annual	SHC
Health Care Practitioners	Does the policy reduce the financial burden for healthcare practitioners?	Press release / social media / policy briefs, etc.	Regular	SHC

Source: Developed by the authors using evaluation findings from the implementation of the Mandatory Medical Malpractice Insurance Policy for Other Health Practitioners (2024), in consultation with the Saudi Health Council.

64 *Magenta Book*, page 81–82.
65 Iskarpatyoti, B. S, et al., page 48.

To effectively engage stakeholders and secure their buy-in, it is advisable to collaboratively develop a usage and dissemination plan. This collaborative approach not only ensures alignment with their expectations but also facilitates broader acceptance and application of the findings.

Academic collaborators may wish to publish evaluation results in scholarly journals.[66] Doing so not only elevates the visibility of the research but also enhances its credibility and disseminates the knowledge more broadly. Encouraging academic publication early in the policy process can also attract a more diverse group of scholars to participate in evaluation activities, enriching the perspectives and expertise involved.

Checklist for Step 6: Use and Disseminate Evaluation Findings	
☐	Establish strategies to ensure that the health policy evaluation findings are effectively utilized.
☐	Maintain ongoing dialogue with the health program team to provide feedback.
☐	Ready stakeholders for the adoption and implementation of evaluation outcomes.
☐	Apply evaluation insights in formulating annual and strategic plans.
☐	Use the findings to advocate for and reinforce the health policy.
☐	Arrange periodic meetings with stakeholders to ensure that the conclusions from the evaluation are clearly communicated and understood
☐	Adapt the presentation of evaluation reports to suit diverse stakeholder needs.
☐	Present the findings clearly and promptly to ensure timely action.
☐	Avoid technical jargon to ensure clarity and accessibility in communications.
☐	Disseminate the evaluation findings through various methods to maximize reach and impact

66 *Magenta Book*, page 82.

References

Ahn, E., and Hyun Kang. 2018. "Introduction to systematic review and meta-analysis." *Korean Journal of Anesthesiology* 71 (2): 103–112. https://doi.org/10.4097/kjae.2018.71.2.103.

American Evaluation Association. 2018. "Guiding Principles for Evaluators." Retrieved from https://www.eval.org/About/Guiding-Principles.

Austin, Elizabeth J., Elsa S. Briggs, Lori Ferro et al. 2023. "Integrating Routine Screening for Opioid Use Disorder into Primary Care Settings: Experiences from a National Cohort of Clinics." *Journal of General Internal Medicine* 38 (2): 332–40. https://doi.org/10.1007/s11606-022-07675-2.

Austin, P. C. 2011. "An introduction to propensity score methods for reducing the effects of confounding in observational studies." *Multivariate Behavioral Research* 46 (3): 399–424. https://doi.org/10.1080/00273171.2011.568786.

Beach, D., and Rasmus B. Pedersen. 2019 (2nd edition). *Process-Tracing Methods: Foundations and Guidelines*. Ann Arbor, USA: University of Michigan Press.

Befani, B., and Gavin Stedman-Bryce. 2017. "Process Tracing and Bayesian Updating for Impact Evaluation." *Evaluation* 23 (1): 42–60. https://doi.org/10.1177/1356389016654584.

Befani, B., Stefano D'Errico, Francesca Booker, and Alessandra Giuliani. 2016. "Clearing the fog: new tools for improving the credibility of impact claims." International Institute for Environment and Development *(IIED) Briefing Papers*. https://www.iied.org/17359iied.

Bonander, C., David Humphreys, and Michelle Degli Esposti. 2021. "Synthetic control methods for the evaluation of single-unit interventions in epidemiology: A tutorial." *American Journal of Epidemiology* 190 (12): 2700–2711. https://doi.org/10.1093/aje/kwab211.

Breckon, J., Sandy Oliver, Cecilia Vindrola, and Thomas Moniz. 2023. *Rapid Evidence Assessments: A guide for commissioners, funders, and policymakers*. CAPE, University College London. https://www.cape.ac.uk/2023/10/31/commissioning-rapid-evidence-assessments/.

Burger, C., Ronelle Burger, and Eddy van Doorslaer. 2022. "The health impact of free access to antiretroviral therapy in South Africa." *Social Science & Medicine* 299: 114832. https://doi.org/10.1016/j.socscimed.2022.114832.

Centers for Disease Control and Prevention. 2013. *Developing an Effective Evaluation Report: Setting the course for effective program evaluation*. Retrieved from https://www.cdc.gov/tobacco/stateandcommunity/tobacco-control/pdfs/developing_evaluation_report.pdf.

Centers for Disease Control and Prevention. 2024. *CDC's Program Evaluation Framework Action Guide*. Atlanta, GA: Centers for Disease Control and Prevention. https://www.cdc.gov/evaluation/media/pdfs/2024/12/FINAL-Action-Guide-for-DFE-12182024_1.pdf.

Centers for Disease Control and Prevention. 2024. "Economic evaluation: Overview." POLARIS. https://www.cdc.gov/polaris/php/economics/index.html.

Collins, A., Deborah Coughlin, James Miller, and Stuart Kirk. 2015. *The Production of Quick Scoping Reviews and Rapid Evidence Assessments: A How to Guide*. London: DEFRA. https://assets.publishing.service.gov.uk/media/5a7f3a76ed915d74e33f5206/Production_of_quick_scoping_reviews_and_rapid_evidence_assessments.pdf.

Crawford, C., Courtney Boyd, Shamini Jain, Raheleh Khorsan, and Wayne Jonas. 2015. "Rapid Evidence Assessment of the Literature (REAL©): Streamlining the systematic review process and creating utility for evidence-based health care." *BMC Research Notes* 8: 631. https://doi.org/10.1186/s13104-015-1604-z.

Davies, R., and Jess Dart. 2005. *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. https://www.mande.co.uk/wp-content/uploads/2005/MSCGuide.pdf.

DeSalvo, K. B., Y. Claire Wang, Andrea Harris, John Auerbach, Denise Koo, and Patrick O'Carroll. 2017. "Public Health 3.0: A Call to Action for Public Health to Meet the Challenges of the 21st century." *Preventing Chronic Disease* 13: E86. https://doi.org/10.5888/pcd13.160017. https://dx.doi.org/10.5888/pcd14.170017.

Dillman, D. A., Jolene D. Smyth, and Leah M. Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, N.J., USA: John Wiley & Sons, Inc.

Ditkowsky, J., Khushal H. Shah, Margaret R. Hammerschlag, Stephan Kohlhoff, and Tamar A. Smith-Norowitz. 2017. "Cost-benefit analysis of *Chlamydia trachomatis* screening in pregnant women in a high burden setting in the United States." *BMC Infectious Diseases* 17 (1): 155. https://doi.org/10.1186/s12879-017-2248-5.

Görgens, M., and Jody Zall Kusek. 2009. *Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit*. Washington, DC: World Bank. https://hdl.handle.net/10986/2702.

Grant, A., Carol Bugge, and Mary Wells. 2020. "Designing process evaluations using case study to explore the context of complex interventions evaluated in trials." *Trials* 21: 982. https://doi.org/10.1186/s13063-020-04880-4.

Hammersley, M., and Paul Atkinson. 2019 (4th edition). *Ethnography: Principles in Practice*. Routledge.

HM Revenue & Customs. 2024. "Soft Drinks Industry Levy: Detailed Information." London: GOV.UK. https://www.gov.uk/government/collections/soft-drinks-industry-levy-detailed-information.

HM Treasury. 2018. "Soft Drinks Industry Levy Comes into Effect." London: GOV.UK. https://www.gov.uk/government/news/soft-drinks-industry-levy-comes-into-effect.

HM Treasury. 2020. *Magenta Book: Central Government Guidance on Evaluation*. Retrieved from https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT_Magenta_Book.pdf.

HM Treasury and Department of Health. 2016. *Soft Drinks Industry Levy: Consultation on the Design of the Levy*. London: HM Treasury. https://assets.publishing.service.gov.uk/media/5a80a0b040f0b62302694998/Soft_Drinks_Industry_Levy-consultation.pdf.

Independent Evaluation Group. 2010. *Cost-benefit analysis in World Bank projects*. Washington, DC: World Bank. https://hdl.handle.net/10986/2561.

Iskarpatyati, B. S., Beth Sutherland, and Heidi W. Reynolds. 2017. *Getting to an Evaluation Plan: A Six-Step Process from Engagement to Evidence*. Retrieved from https://www.measureevaluation.org/resources/publications/ms-17-124/at_download/document.

Iwashyna, T. J., and Edward H. Kennedy. 2013. "Instrumental variable analyses: Exploiting natural randomness to understand causal mechanisms." *Annals of the American Thoracic Society* 10 (3): 255–260. https://doi.org/10.1513/AnnalsATS.201303-054FR.

Jack, S. M., Andrea Gonzalez, Karen Campbell, et al. 2020. *Implementation and delivery of Nurse-Family Partnership in British Columbia, Canada: A synthesis of selected findings from the British Columbia Healthy Connections Project Process Evaluation (2013–2018).* Hamilton, ON, Canada: School of Nursing, McMaster University. https://phnprep.ca/wp-content/uploads/2021/09/BCHCP_Process-Evaluation_Final-Report.pdf.

Joint Committee on Standards for Educational Evaluation. 2016. "The program evaluation standards" (3rd ed.). Ypsilanti, MI: Western Michigan University. Retrieved from https://files.wmich.edu/s3fs-public/attachments/u350/2021/program-eval-standards-jc.pdf.

Kim, D., and Kim, M. 2022. Impact of the COVID-19 Pandemic on Mental Health Services in Korea: Challenges and Opportunities. *Journal of Preventive Medicine & Public Health*, 55(4), 195–202. https://doi.org/10.3961/jpmph.22.195.

Kingdom of Saudi Arabia Vision 2030, *Health Sector Transformation Program*. https://www.vision2030.gov.sa/en/explore/programs/health-sector-transformation-program.

Kingdom of Saudi Arabia. 2020-2021. "Health Sector Transformation Program Delivery Plan." https://www.vision2030.gov.sa/media/u5xapka3/2021-2025-health-sector-transformation-program-delivery-plan-en.pdf.

Lopez Bernal, J., Steven Cummins, and Antonio Gasparrini. 2017. "Interrupted time series regression for the evaluation of public health interventions: A tutorial." *International Journal of Epidemiology* 46 (1): 348–355. https://doi.org/10.1093/ije/dyw098.

Mayne, J. 2001. "Addressing attribution through contribution analysis: Using performance measures sensibly." *Canadian Journal of Program Evaluation* 16 (1): 1–24. https://doi.org/10.3138/cjpe.016.001.

Mayne, J. 2012. "Making causal claims." Institutional Learning and Change Initiative (ILAC) Brief 26. https://hdl.handle.net/10568/70211.

National Academies of Sciences, Engineering, and Medicine. 2023. *Review of four CARA programs and preparing for future evaluations*. Washington, DC: The National Academies Press. https://doi. org/10.17226/26831.

National Committee of Bioethics (NCBE). "About the NCBE." King Abdulaziz City for Science and Technology (KACST). Retrieved from https://ncbe.kacst.edu.sa/en/about-us/who-we-are/.

OECD (Organisation for Economic Co-operation and Development). 2019. "Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use." Paris: OECD Publishing. https://doi.org/10.1787/15a9c26b-en.

OECD. 2020. *Improving Governance with Policy Evaluation: Lessons From Country Experiences*. Paris: OECD Publishing. Retrieved from https://www.oecd-ilibrary.org/governance/improving-governance-with-policy-evaluation_89b1577d-en.

OECD 2020. Regulatory Impact Assessment, OECD Best Practice Principles for Regulatory Policy. Paris: OECD Publishing. https://doi.org/10.1787/7a9638cb-en.

Office for Health Improvement and Disparities. 2018. "Outcome evaluation: evaluation in health and wellbeing." GOV.UK. Department of Health and Social Care. https://www.gov.uk/guidance/evaluation-in-health-and-wellbeing-outcome.

Pan American Health Organization (PAHO). 2013. *Health Impact Assessment: Concepts and Guidelines for the Americas*. Washington, DC: PAHO. Retrieved from https://www3.paho.org/hq/dmdocuments/2014/health-impact-assessment-concepts-and-guidelines-2013.pdf.

Papadakis, S., Adam G. Cole, Robert D. Reid, et al. 2016. "Increasing rates of tobacco treatment delivery in primary care practice: Evaluation of the Ottawa Model for Smoking Cessation." *Annals of Family Medicine* 14 (3): 235-43. https://doi.org/10.1370/afm.1909.

Patton, M. Q. 2015. *Qualitative Research and Evaluation Methods*. 4th ed. Thousand Oaks, CA, USA: SAGE Publications, Inc.

Pawson, R., and Nicholas Tilley. 1997. *Realistic Evaluation*. SAGE Publications.

Ragin, C. C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Oakland, CA, USA: University of California Press.

Sasabuchi, Y. 2022. "Introduction to regression discontinuity design." *Annals of Clinical Epidemiology* 4 (1): 1–5. https://doi.org/10.37737/ace.22001.

Sattar, R., Rebecca Lawton, Maria Panagioti, and Judith Johnson. 2021. "Meta-ethnography in healthcare research: A guide to using a meta-ethnographic approach for literature synthesis." BMC Health Services Research 21 (1): 50. https://doi.org/10.1186/s12913-020-06049-w.

Saudi Data and Artificial Intelligence Authority (SDAIA). 2021. *Personal Data Protection Law*. https://sdaia.gov.sa/en/SDAIA/about/Documents/Personal%20Data%20English%20V2-23April2023-%20Reviewed-.pdf.

Saudi Data and Artificial Intelligence Authority. 2022. *Personal Data Protection Law*. https://sdaia.gov.sa/ar/SDAIA/about/Files/RegulationsAndPolicies02.pdf.

Saudi Health Council. 2024. Evaluate the Implementation of Mandatory Medical Malpractice Insurance for Other Health Practitioners Policy.

Saudi Health Council and World Bank. 2024. *The Health Policy Maker's Manual: Integrating Data and Evidence*. https://shc.gov.sa/Arabic/Documents/The%20Health%20Policy%20Makers%20Manual%20-%20KSA%20-2024.pdf.

Schick-Makaroff, K., Marjorie MacDonald, Marilyn Plummer, Judy Burgess, and Wendy Neander. 2016. "What synthesis methodology should I use? A review and analysis of approaches to research synthesis." *AIMS Public Health* 3 (1): 172–215. https://doi.org/10.3934/publichealth.2016.1.172.

Schneider, C. Q., and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge, U. K: Cambridge University Press.

Sheingold, S. H., and Anupa U. Bir. (2019). *Evaluation for Health Policy and Health Care: A Contemporary Data-Driven Approach*. SAGE Publications. Retrieved from https://us.sagepub.com/hi/nam/evaluation-for-health-policy-and-health-care/book262472.

Simms, K. T., Jean-François Laprise, Megan A. Smith et al. 2016. "Cost-effectiveness of the next generation nonavalent human papillomavirus vaccine in the context of primary human papillomavirus screening in Australia: a comparative modelling analysis." *Lancet Public Health* 1 (2): E66–E75. https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(16)30019-6/fulltext.

Thomas, J., Mark Newman, and Sandy Oliver. 2013. "Rapid evidence assessments of research to inform social policy: taking stock and moving forward." *Evidence and Policy* 9 (1): 5–27. https://doi.org/10.1332/174426413X662572.

UK Government Social Research Service. 2014. *Rapid Evidence Assessment Toolkit*. https://webarchive.nationalarchives.gov.uk/ukgwa/20140402164155/http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment.

UNAIDS. n.d. *Indicator Standards: Operational Guidelines for Selecting Indicators for the HIV Response*. Retrieved from https://www.globalhivmeinfo.org/AgencySites/MERG%20Resources/MERG%20Indicator%20Standards_Operational%20Guidelines.pdf.

U.S. Department of Health and Human Services (HHS). 2016. *Guideline for Regulatory Impact Analysis*. https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//171981/HHS_RIAGuidance.pdf.

Virani, S. S., Alvaro Alonso, Hugo J. Aparicio et al. 2021. "Heart disease and stroke statistics—2021 update: A report from the American Heart Association." *Circulation* 143 (8): e254-e743. https://doi.org/10.1161/CIR.0000000000000950.

White, H., and David A. Raitzer. 2017. *Impact Evaluation of Development Interventions: A Practical Guide*. Metro Manila, Philippines: Asian Development Bank. Retrieved from https://www.adb.org/sites/default/files/publication/392376/impact-evaluation-development-interventions-guide.pdf.

Wilson-Grau, R., and Heather Britt, H. 2012. *Outcome Harvesting*. Ford Foundation. https://outcomeharvesting.net/wp-content/uploads/2016/07/Outcome-Harvesting-Brief-revised-Nov-2013.pdf.

Wing, C., Kosali Simon, and Ricardo A. Bello-Gomez. 2018. "Designing difference in difference studies: Best practices for public health policy research." *Annual Review of Public Health* 39: 453–469. https://doi.org/10.1146/annurev-publhealth-040617-013507.

World Bank. *Cost-effectiveness analysis*. DIME Wiki. https://dimewiki.worldbank.org/Cost-effectiveness_Analysis.

World Bank. 2004. Monitoring and Evaluation: Some Tools, Methods, and Approaches (Updated Edition). Washington, DC: World Bank. https://hdl.handle.net/10986/23975.

World Bank. 2005. *The Logical Framework (Logframe) Handbook: A Logical Framework Approach to Project Cycle Management*. https://documents1.worldbank.org/curated/en/783001468134383368/pdf/31240b0LFhandbook.pdf.

World Bank. 2019. "World Bank Group Evaluation Principles." Retrieved from https://ieg.worldbankgroup.org/sites/default/files/Data/reports/WorldBankEvaluationPrinciples.pdf.

World Bank. 2020. *World Development Report 2020: Trading for Development in the Age of Global Value Chains*. Washington, DC: World Bank. https://hdl.handle.net/10986/32437.

Yin, R. K. 2014. *Case Study Research: Design and Methods*. Thousand Oaks, CA, USA: SAGE Publications, Inc.

Zabor, E. C., Alexander M. Kaizer, and Brian P. Hobbs. 2020. Randomized controlled trials. *Chest* 158 (1 Suppl): S79–S87. https://doi.org/10.1016/j.chest.2020.03.013.

Zhou, F., Abigail Shefer, Jay Wenger, Mark Messonnier, Li Yan Wang, Adriana Lopez, Matthew Moore, Trudy V. Murphy, Margaret Cortese, and Lance Rodewald. 2014. "Economic Evaluation of the Routine Childhood Immunization Program in the United States, 2009." *Pediatrics* 133 (4): 577-585. https://doi.org/10.1542/peds.2013-0698.

08

Annexes

Annex 1: Regulatory Impact Analysis

What is Regulatory Impact Analysis (RIA)?

Regulatory Impact Analysis (RIA)[67] is a structured, ex-ante framework for assessing the anticipated effects of health policy or regulatory options. Its primary goal is to ensure that regulations are justified, necessary, and aligned with public health and social objectives, while being implemented efficiently, equitably, and cost-effectively.

RIA evaluates expected benefits, costs, and trade-offs, considering alternatives and additional impacts as mandated by Saudi Cabinet decisions or Ministry of Health (MoH) guidelines. It promotes transparency for stakeholders, such as the SHC, healthcare providers, and the public, by quantifying and monetizing effects where possible. These effects include estimating the economic and health impacts, and analyzing their distribution across populations. For instance, the RIA of Saudi Arabia's 2017 sugar tax could assess its impact on reducing obesity rates – a key factor in diabetes prevalence – among adults versus children, ensuring equitable health outcomes and efficient resource use in line with Vision 2030 goals.

Why is Regulatory Impact Analysis Important?

Conducting an RIA in health policy evaluation offers significant benefits, though it also entails certain costs.

Key Benefits

1. **Better health policy decisions.** RIA provides a robust evidence base, documenting data, assumptions, and analyses to inform decisions. It offers insights into critical outcomes, even those hard to quantify—like the psychological benefits of expanding mental health services under Saudi Arabia's Vision 2030.
2. **Thorough impact assessment.** RIA identifies intended and unintended impacts – for example, that introducing mandatory health insurance in Saudi Arabia could potentially increase out-of-pocket costs for low-income families. It also clarifies stakeholder views, fostering transparency and trust.
3. **Stronger public health outcomes.** By analyzing benefits and costs, RIA ensures that regulations maximize health gains, enhance equity, and reduce burdens. For example, RIA could assess whether centralizing oncology services in major Saudi cities like Riyadh lowers costs while maintaining equitable access for rural and underserved communities.

Associated Costs

1. **Resource demands.** RIA requires time, expertise, and funding for thorough assessments—such as collecting nationwide data to evaluate the cost-effectiveness of Saudi Arabia's national influenza or HPV vaccination programs.
2. **Tailored analysis.** To manage resources, RIA must prioritize key health outcomes and actionable insights, focusing on critical metrics like disease incidence or hospitalization rates rather than secondary indicators like patient appointment punctuality.

When is a Regulatory Impact Analysis Performed?

RIA is critical when a proposed health policy or regulation is expected to have significant impacts on the economy, public health, or societal well-being. It supports thorough evaluation of these regulations to maximize benefits, minimize unintended consequences, and promote equitable and efficient outcomes.

RIA is particularly important in the following scenarios:

1. **Substantial economic impact.** When a regulation imposes major financial implications for the healthcare sector, RIA assesses whether its anticipated benefits justify the costs.
    - » **For example:** A regulation requiring all hospitals to adopt electronic health records (EHR) within a short timeframe would impose significant costs for technology implementation. RIA would evaluate whether the benefits of improved patient care and data management outweigh these costs, exploring cost-effective alternatives.

67 Guideline for Regulatory Impact Analysis, US Department of Health and Human Services (HHS). https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//171981/HHS_RIAGuidance.pdf.

2. **Broad societal impact.** When a regulation significantly affects public health, safety, or welfare, RIA evaluates its overall impact and feasibility.

    » **For example:** A policy mandating nationwide vaccination for an emerging infectious disease could prevent outbreaks and save lives but might raise concerns about vaccine accessibility and public compliance. To ensure positive societal impact, RIA would analyze public health benefits, economic costs of vaccine procurement and distribution, and potential resistance.

3. **Significant distributional effects.** When a policy is likely to disproportionately affect specific groups, such as low-income populations or rural communities, RIA identifies and addresses equity concerns.

    » **For example:** A policy to increase taxes on sugary beverages to reduce consumption and combat obesity might disproportionately burden lower-income populations who spend a larger share of their income on such products. RIA would examine these impacts to ensure the policy achieves health goals without unfairly affecting vulnerable groups.

4. **Complex or uncertain outcomes.** When a regulation involves complex interventions or has uncertain, far-reaching consequences, RIA provides a framework to evaluate trade-offs and navigate complexities.

    » **For example:** New air quality standards in healthcare facilities could improve patient outcomes by reducing pollution-related illnesses but might increase operational costs. RIA would analyze health benefits alongside economic impacts on facilities, providing evidence to guide policy makers in weighing these trade-offs.

The scope and depth of RIA depend on the regulation's anticipated impact—comprehensive or limited—ensuring efficient resource use while delivering critical insights for decision-making.

- **Comprehensive RIA for significant impact:** For regulations with substantial effects, such as those profoundly impacting public health, the economy, or societal well-being, a detailed analysis is essential.

    » **For example:** Introducing mandatory health insurance for citizens under a universal health coverage goal would require a comprehensive RIA. This would evaluate public health benefits such as improved access to essential healthcare; economic costs like funding and administration; and societal implications such as affordability for low-income groups and impacts on private providers; ensuring that the policy maximizes health outcomes while maintaining financial sustainability and equity.

- **Streamlined RIA for limited impact:** For regulations with minor impacts such as adjustments to existing policies, a streamlined analysis focuses on key aspects, avoiding unnecessary complexity.

    » **For example:** A regulation updating nutritional labeling requirements on food products would likely have a limited impact. A streamlined RIA could evaluate compliance costs for manufacturers and the potential public health benefits of clearer labeling, without requiring extensive analysis.

Whether comprehensive or streamlined, RIA enhances decision-making by providing a clear, evidence-based rationale, ensuring that regulatory actions align with health policy goals like improving public health, enhancing equity, or optimizing healthcare efficiency.

Case Study **Regulatory Impact Analysis of the United Kingdom's Soft Drinks Industry Levy**

Overview: The UK's Soft Drinks Industry Levy (SDIL), implemented in April 2018, exemplifies how Regulatory Impact Analysis (RIA) informs health policy decisions. Known as the sugar tax, SDIL targets soft drinks with added sugar to reduce consumption, combat obesity, and address related issues like type 2 diabetes and dental caries. This case study highlights the RIA conducted in 2016 by HM Treasury and the Department of Health, demonstrating its methodology, findings, and initial impact, to provide a model for similar evaluations, including in Saudi Arabia under Vision 2030.

Policy Context: Announced in March 2016 and launched in April 2018, the SDIL levies a two-tiered tax on soft drinks: £0.24 per liter for drinks with 8g or more of sugar per 100ml (high tier); £0.18 per liter for drinks with 5–8g per 100ml (low tier); and exemptions for drinks with less than 5g sugar per 100ml, fruit juices, milk-based drinks, and small producers (less than 1 million liters annually). The policy aimed to encourage manufacturers to reduce sugar content, promote low-sugar alternatives, and decrease portion sizes, addressing the UK's obesity epidemic, where rates have nearly doubled over 30 years.

Objectives of the RIA: The 2016 RIA aimed to:

- Assess potential health benefits from reduced sugar consumption, particularly lowering obesity, type 2 diabetes, and dental caries.
- Evaluate economic impacts, including costs to the soft drinks industry – for example, reformulation and compliance, and government revenue generation.
- Examine distributional effects to ensure that the tax does not disproportionately burden specific income groups, promoting equity.

These objectives reflect RIA's role as an ex-ante, evidence-based tool to inform regulatory decisions, balancing health, economic, and social outcomes.

Methodology: The RIA used a mixed-methods approach, including:

- **Health impact assessment:** Epidemiological models predicted sugar consumption reductions and potential health outcomes, estimating fewer obesity and disease cases based on national health surveys and consumption data.
- **Economic analysis:** Assessed industry costs, such as one-off familiarization and ongoing compliance, and projected government revenue (approximately £520 million annually) using cost-benefit analysis and consumption forecasts, accounting for reformulation and price changes.
- **Stakeholder consultation:** Conducted a public consultation in 2016, engaging industry – manufacturers, retailers – public health bodies, and consumers to address concerns including costs, health priorities, and equity, ensuring transparency.
- **Distributional analysis:** Analyzed consumption patterns across income groups, finding minimal risk of disproportionate burden on lower-income households, supported by equity impact assessments.

Key Findings:

- **Health impacts.** Predicted significant reductions in sugar consumption, potentially preventing cases of obesity, type 2 diabetes, and dental caries over time. Post-implementation data (up to 2020) confirmed a 46% average sugar content reduction in levy-eligible drinks (2015–2020) and over 45,000 tons of sugar removed annually, aligning with initial projections.
- **Economic impacts.** Projected £520 million in annual revenue, with costs primarily borne by manufacturers through reformulation and compliance. Industry faced one-off and ongoing costs, but the impact on the estimated 300 UK producers was expected to be negligible, with potential price pass-through to consumers noted.
- **Distributional effects.** Found the tax would have a relatively even impact across income groups, minimizing equity concerns. But the tax had minimal effects on specific health conditions such as type 1 diabetes and lactose intolerance; these were acknowledged and mitigated by exemptions.

Note: The £520 million revenue projection highlighted the policy's potential to fund public health initiatives while improving health, though 2018 data revised this to £240 million annually due to reformulation, indicating initial overestimation.

Conclusion and implications: The RIA supported the SDIL's implementation, finding that health benefits likely outweighed costs. Initial post-implementation outcomes (2018–2020) validated sugar reductions and revenue generation, but uncertainties—consumer behavior shifts, long-term health impacts, industry responses—were noted, emphasizing ongoing evaluation. This demonstrates RIA's value in data-driven policymaking, ensuring regulations are justified, efficient, and equitable.

Sources: HM Treasury and Department of Health. 2016. *Soft Drinks Industry Levy: Consultation on the Design of the Levy* https://assets.publishing.service.gov.uk/media/5a80a0b040f0b62302694998/Soft_Drinks_Industry_Levy-consultation.pdf; HM Revenue & Customs. 2024. *Soft Drinks Industry Levy: Detailed Information*. https://www.gov.uk/government/collections/soft-drinks-industry-levy-detailed-information; HM Revenue & Customs. 2018. *Soft Drinks Industry Levy Comes into Effect*. https://www.gov.uk/government/news/soft-drinks-industry-levy-comes-into-effect.

Annex 2: Rapid Evidence Assessment

Rapid Evidence Assessment

Quick review for health policy evaluation

Definition

A Rapid Evidence Assessment (REA) is a streamlined, yet systematic method of reviewing existing evidence to inform health policy decisions within a limited timeframe. In the context of health policy evaluation, REAs are used to quickly assess what is already known about the effectiveness, implementation, cost-effectiveness, or unintended consequences of health interventions or programs. While not as exhaustive as full systematic reviews, REAs apply rigorous and transparent processes to ensure the evidence collected is relevant, credible, and useful for timely policy-making.

When to use REAs?

This approach is particularly valuable in dynamic healthcare environments where decisions must be informed by the best available evidence, but time and resources are constrained. By focusing on targeted research questions and using defined inclusion criteria, REAs support policy makers in making evidence-informed decisions.

Benefits of Rapid Evidence Assessment

- **Timely:** REAs are designed to deliver findings much faster than traditional systematic reviews—typically within weeks to a few months. This makes them especially useful in policy environments where decisions must be made quickly.
- **Systematic:** Although abbreviated, REAs still apply a transparent and replicable methodology, including predefined criteria for study selection and data synthesis. This ensures that the process is methodical and evidence-based.
- **Cost-effective:** By streamlining certain steps of the full systematic review – for example, limiting databases, or focusing on recent studies – REAs reduce the time and labor involved, resulting in lower costs.
- **Flexible:** REAs can be tailored to address a wide range of policy questions. The scope, depth, and sources can be adjusted according to the specific needs and resources available.

Limitations of Rapid Evidence Assessment

- **Reduced comprehensiveness:** To meet time constraints, REAs often limit the number of databases searched or the years of publication reviewed. This may result in missing some relevant studies.
- **Potential bias:** Due to shortened timelines, REAs may exclude gray literature, increasing the risk of publication bias.
- **Limited depth:** REAs typically provide high-level summaries of study findings rather than detailed analysis or in-depth synthesis, which can limit their utility for complex decision-making.
- **Less suitable for complex questions:** REAs work best for relatively straightforward questions. They may not be appropriate for evaluating complex interventions, systems, or theories of change that require in-depth qualitative or mixed-method analysis.

Steps to Conduct a Rapid Evidence Assessment

1. **Define the policy question and scope**

   Clearly articulate the specific health policy question the REA is intended to answer. This may relate to the effectiveness, efficiency, equity, or implementation of a health policy. The scope should focus on the evidence needed to inform timely policy decisions.

2. **Develop a brief protocol**

   Prepare a short protocol outlining the objectives, inclusion and exclusion criteria, search strategy, health outcomes of interest, and policy relevance. The protocol should align with the needs of policy makers and ensure a consistent and transparent approach.

3. **Conduct a targeted literature search**

   Search relevant health databases such as PubMed, Cochrane Library, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL), along with gray literature from trusted sources—for example, the World Health Organization (WHO) and national health authorities. The search should be focused on recent and high-quality evidence directly relevant to the policy issue.

4. **Screen and select relevant studies**

   Review titles, abstracts, and full texts based on predefined criteria. Select studies that are methodologically sound and policy-relevant, such as those evaluating real-world effectiveness or cost-effectiveness in healthcare settings.

5. **Extract and synthesize health policy-relevant data**

   Extract key findings such as intervention type, population characteristics, outcomes – for example, reduced morbidity, healthcare utilization – and policy implications. A synthesis is typically narrative but can include tables, rating systems, or logic models to visualize health impacts.

6. **Report findings and implications for health policy**

   Provide a clear and concise summary of the evidence, emphasizing its relevance to current health policy decisions. Discuss the strength of the evidence, identify any limitations, and offer actionable recommendations where appropriate.

Sources: Adapted from UK Government Social Research Service. 2014. *Rapid Evidence Assessment Toolkit*; Thomas, J., Newman, M., & Oliver, S. 2017. *Rapid evidence assessments of research to inform social policy: taking stock and moving forward*; Collins, A., Coughlin, D., Miller, J., & Kirk, S. 2015. *The Production of Quick Scoping Reviews and Rapid Evidence Assessments: A How to Guide*. DEFRA; Breckon, J., Oliver, S., Vindrola, C., and Moniz, T. 2023. Rapid Evidence Assessments: A guide for commissioners, funders, and policymakers. https://www.cape.ac.uk/2023/10/31/commissioning-rapid-evidence-assessments/.

Annex 3: Key Questions for Each Evaluation Type (Non-Exhaustive)

Formative Evaluation Questions	Process Evaluation Questions	Outcome Evaluation Questions	Impact Evaluation Questions	Economic Evaluation Questions
How is the health policy intended to be implemented across relevant settings – for example, facilities, regions, or sectors?	How is the policy being delivered? Are there any challenges or deviations from the implementation as planned?	Did the target population receive the intended services or products?	Does the health policy work? What are its positive and negative effects, intended and unintended?	Is the health policy cost-effective? Does it provide good value for the resources invested?
What data are currently being collected, and what additional information is needed to assess the health policy's feasibility and readiness for implementation?	What implementation methods and strategies are being used?	Did the target population, such as specific age groups or underserved communities, receive the intended health services or products, such as vaccinations or screenings?	Does the health policy achieve its long-term goals, such as improved population health or reduced disparities?	Is the health policy cost-effective in achieving its health outcomes compared to planned targets or benchmarks?
To what extent are the identified target groups being considered and adequately addressed in the policy design?	What challenges or barriers are encountered in the implementation process?	What specific health benefits or outcomes were achieved, such as increased vaccination rates or reduced hospital admissions?	What are the positive intended effects of the policy, such as better access to care or reduced chronic disease rates?	Does the policy provide good value for the resources invested, such as financial, human, and time resources?

Formative Evaluation Questions	Process Evaluation Questions	Outcome Evaluation Questions	Impact Evaluation Questions	Economic Evaluation Questions
How well are the proposed policy components aligned with the needs, expectations, and capacities of the target groups and implementing actors?	Have there been any deviations from the planned implementation approach, and what are the reasons?	Was the policy efficient in delivering benefits, for example, in terms of speed, resource use, and reach, compared to expectations or alternatives?	What are the negative unintended effects, if any, such as increased healthcare disparities or unintended burdens on providers?	What are the costs of achieving the policy's health outcomes, such as reduced morbidity or improved quality of life?
What factors are likely to facilitate or hinder the successful implementation of the health policy in its intended context?	To what extent has the policy reached its intended target populations—for example, rural areas or low-income groups.		To what extent can the observed effects be attributed to the policy rather than external factors, such as economic conditions or other programs?	What are the implementation costs associated with the policy—for example, training, infrastructure, and personnel?
What modifications to the policy design or delivery approach could improve its relevance, acceptability, or operational feasibility before scale-up?	How do external factors, such as pandemics or economic shifts, influence the implementation process?		What causal factors, such as policy design or community engagement, contributed to the policy's effects?	Has the policy been cost-effective compared to alternative interventions or no intervention?
	How do contextual factors, such as cultural, social, or environmental factors, affect the implementation of the policy?		Has the policy resulted in any unintended long-term health or social outcomes?	What is the value-for-money of the health policy, such as the cost-benefit ratio?
	How have the attitudes and behaviors of stakeholders—for example, populations, providers—influenced the implementation process?		Have the impacts been influenced by other external factors, such as technological advances?	What are the benefits of the policy, such as improved quality of life or reduced healthcare costs?
	What areas could be improved in the current implementation strategies to enhance effectiveness or efficiency?		To what extent have different population groups, such as by age, income, or region, been impacted differently, and why?	What are the costs of implementing and sustaining the policy, such as financial, time, and workforce costs?
			Can the policy's health impacts be reproduced or scaled in other regions or contexts?	Do the benefits outweigh the costs, and how does this compare to alternative health policies?
				What is the ratio of costs to benefits, and how does it compare to other health policy options?

Annex 4: Sample Data Collection Plan

#	Indicator	Data Needed	Data Collection Method/Source	From Whom Will These Data Be Collected	By Whom and When Will Data Be Collected	Security or Confidentiality Steps
1	% of patients' allegations against healthcare practitioners for negligence	Total number of allegations against practitioners; total number of patients served	SHC legal reporting system and service statistics	Healthcare facilities and SHC records	SHC Legal Affairs Unit, annually	Facility-level data, anonymized before analysis
2						
3						
4						
5						

Annex 5: Prioritized Health Status and Health System Indicators

Health Status Indicators - 51 Indicators			**Sustainable Development Goal (SDG)**
Topic	**#**	**Indicator Name**	
Awareness & Education	1	Completion rate (primary education, lower secondary education, upper secondary education)	SDG 4.1.2
	2	Health Literacy rate	
Behavioral	3	Prevalence of smoking	
	4	Physical activity	
	5	Population (15 years and over), by number of eating fruits and vegetables servings per day and Age Group	
Child Development	6	Proportion of children under 5 years of age who are developmentally on track in health, learning, and psychosocial well-being	SDG 4.2.1
Demographics	7	Percentage of people age 65 and older	
	8	Population by sex/age	
	9	Population growth	
	10	Total Population (Millions of people)	
Environmental Factors	11	Air pollution level in cities (particulate matter [PM2.5])	
	12	Proportion of population using safely managed drinking-water services	
	13	Proportion of population using safely managed sanitation services	SDG 6.2.1
Function	14	Children under 5 years who are overweight	SDG 2.2.2
	15	Incidence of low birth weight among newborns	
	16	Age-standardized prevalence of overweight and obesity in persons aged 18+ years	
	17	Age-standardized prevalence of raised blood pressure among persons aged 18+ years	
Socio-economic	18	Proportion of population living below the national poverty line, by sex and age	SDG 1.2.1
	19	Unemployment rate, by sex, age and persons with disabilities	SDG 8.5.2

Health Status Indicators - 51 Indicators			**Sustainable Development Goal (SDG)**
Topic	**#**	**Indicator Name**	
Care Coverage	20	Percentage of women aged 15–49 who received four or more antenatal care visits	
	21	Births attended by skilled healthcare personnel	SDG 3.1.2.
Immunization	22	Immunization coverage rate by vaccine for each vaccine in the national schedule	SDG 3.b.1
	23	Percentage of targeted population vaccinated with seasonal influenza vaccine	
	24	Percentage of specific communicable disease that achieved targeted decrease	
Program	25	Number of National Prevention Programs for NCDs	
	26	Percentage improvement resulting from the National Prevention Program	
Screening	27	% of Primary Health Care Center (PHCC) visitors (18 years and above) screened for hypertension	
	28	% of PHCC visitors (2 years and above) screened for obesity and overweight	
	29	% of PHCC visitors (40 years and above) screened for diabetes	
	30	% of PHCC visitors (40 years and above) screened for dyslipidemia	
	31	% of PHCC visitors aged 50 years and above screened for colorectal cancer by fecal immunochemical test (FIT)	
	32	% of women aged 40-69 years screened for breast cancer using mammogram	
	33	% of PHCC visitors (adults) screened for COPD (Chronic Obstructive Pulmonary Disorder)	
Mental Well-being	34	Proportion of adults with psychological distress	
Incidence	35	Cancer incidence rate, by type of cancer (per 100, 000 population)	
	36	Incidence of heart attacks (acute coronary events)	
Incidence (Incidence of Communicable Diseases)	37.1	Estimated number of new hepatitis B infections per 100,000 population in a given year	Hepatitis B: SDG 3.3.4
	37.2	HIV incidence (per 1000 population)	HIV: SDG 3.3.1
	37.3	Incidence of meningococcal meningitis in KSA	
	37.4	Incidence rate of hepatitis C virus (HCV)	
	37.5	Incidence rate of measles	
	37.6	Incidence rate of rubella	
	37.7	Malaria incidence rate (per 1,000 population)	Malaria: SDG 3.3.3
	37.8	TB incidence (per 100 000 population)	TB: SDG 3.3.2
Injuries	38	The number of serious injuries resulting from traffic accidents per 100,000 population	
Prevalence	39	Population (15 years and above) who suffer from A Chronic Disease by Name of Diagnosed Disease	
Life Expectancy	40	Health expectancy: Healthy Life Years (HLY)	
	41	Life expectancy at birth	
Mortality by Age	42	Mortality Rate	

Health Status Indicators - 51 Indicators			Sustainable Development Goal (SDG)
Topic	**#**	**Indicator Name**	
Mortality by Cause - Chronic Condition	43	Mortality rate due to cardiovascular disease, cancer, diabetes, and chronic respiratory diseases	
Mortality by Cause – Environmental: Air Quality	44	Mortality attributable to joint effects of household and ambient air pollution	SDG 3.9.1
Mortality by Cause – Environmental: Water	45	Mortality rate attributed to unsafe water, unsafe sanitation, and lack of hygiene (exposure to unsafe water, sanitation, and hygiene for all (WASH) services)	SDG 3.9.2
Mortality by Cause – Maternity	46	Maternal mortality ratio (per 100,000 live births)	SDG 3.1.1
Mortality by Cause – Premature NCDs	47	Premature noncommunicable disease (NCD) mortality	SDG 3.4.1
Mortality by Cause – Suicide	48	Suicide rate (per 100,000 population)	SDG 3.4.2
Mortality by Cause – Traffic Accidents	49	Death rate due to road traffic injuries	SDG 3.6.1
Mortality by Cause – Communicable Diseases	50.1	Dengue mortality rate	
	50.2	Malaria mortality rate (per 100,000 population)	
	50.3	Tuberculosis (TB) mortality rate (per 100,000 population)	
Mortality by Cause	51	COVID-19 Cases: Deaths	

Health System Indicators - 58 Indicators			Sustainable Development Goal (SDG)
Topic	**#**	**Indicator Name**	
Functionality	1	Functional Health Outcomes Score: Inpatient and Outpatient	
Mortality	2	Avoidable mortality (preventable and treatable)	
Perception of Health	3	Population (15 years and above) who assessed their own health status as good or very good	
Quality of Life	4	Quality of Life	
Care Coordination	5	Experienced a coordination problem in the past two years	
	6	Patient assessment of level of integration in health care delivery	
	7	Post-care encounter	
Communication	8	Care from clinicians (doctors and nurses)	
Overall Rating	9	% Timely resolution of complaints from patients and their families	
	10	Overall hospital rating	
	11	Patient experience score	
Shared Decision Making	12	Access to own medical record	
	13	Families feeling involved in the care of the patient	
	14	Patients feeling involved in the decision making of their care	
Workforce Experience	15	Working life experience	
Care Coverage	16	Unmet health care needs	

Health System Indicators - 58 Indicators			Sustainable Development Goal (SDG)
Topic	**#**	**Indicator Name**	
Technology	17	Percentage of health centers implementing the electronic Health Information Systems (HIS)	
	18	Percentage of residents who have a unified digital medical record	
Time to Care	19	Days to third next available appointment	
	20	Emergency Department (ED) Median Time from ED arrival to ED departure for discharged ED patients for Adult Patients	
	21	Waiting times for elective surgery: proportion admitted within clinically recommended time	
	22	Outpatient Department (OPD) Appointment Waiting Time (Days) (OPD including specialist)	
	23	Average number of days a patient waits for an appointment (Urgent referral)	
Proximity to Care	24	Percentage of polyclinics in PHCCs that have three or more of the main specialties (medicine, obstetrics & gynecology, surgery, and pediatrics) of polyclinics per region	
Utilization	25	Number of Primary Health Care (PHC) visits per capita per year	
	26	Outpatient service utilization	
	27	Patient seen in Virtual clinic	
Evidence- based practice	28	Low value interventions	
	29	Preventable hospitalization rate	
Maternity	30	Births by caesarean section (%)	
	31	Patients with elective vaginal deliveries or elective cesarean births at ≥ 37 and < 39 weeks of gestation completed	
Condition- specific Effectiveness	32	Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9/0%)	
	33	Thrombolytic Therapy	
Compliance	34	Percentage of hospitals that comply with Central Board for Accreditation of Healthcare Institutions (CBAHI) Essential National Requirements (ESR) requirement	
	35	Percentage of hospitals that met US median of patient safety culture survey	
Events	36	Rate of Adverse Events	
	37	Rate of Sentinel Events	
Infection	38	Hospital acquired infection rate	
Readmission	39	Hospital Readmission Percentage	
Readmission (by department)	40	Percentage (%) of all patients who re-attend the emergency room within 72 hours (3 days) of a previous visit to the ER	
Availability of Medication	41	% Of the basic medicines available in the local market	
Beds	42	Bed Management / Bed Occupancy (%)	
	43	Hospital bed density (per 10,000 population)	
Cost	44	Cost per weighted separation and total case weighted separations	
Density	45	Health facility density and distribution	
	46	Health worker density and distribution	SDG 3.c.1
Hospital Stay	47	Average length of hospital stay (in days)	

Health System Indicators - 58 Indicators			Sustainable Development Goal (SDG)
Topic	**#**	**Indicator Name**	
Spend	48	General government expenditure on health as a percentage of total government expenditure (ten-year growth)	
	49	Healthcare expenditure per capita	
	50	Out-of-pocket expenditure as percentage of current health expenditure (CHE)	
	51	Population covered by Health Insurance	
	52	Public domestic sources of current spending on health as % of current health expenditure	
	53	Private domestic sources of current spending on health as % of current health expenditure	
	54	Spend of low value interventions	
	55	Total current expenditure on health as percentage of gross domestic product	
	56	Total revenue generated from private health insurance and out-of-pocket spend for utilizing government health resources	The definition and methodology of this indicator was adjusted for simplification
Workforce	57	Net growth in health workforce	
	58	Saudization (%)	

Annex 6: Key Assessment Points to Validate Indicators

The newly developed indicators could be generated following the below key criteria:

Criteria 1: The Indicator Should Be Essential and Practical

Criteria 1: The indicator should be essential and practical		
	Assessment Points	**Explanation for Each Assessment Point**
☐	Justified	The necessity of the indicator in evaluating the health policy should be justified.
☐	Valuable and relevant	Evaluators should identify which stakeholders will use the indicator. To ensure that the indicator is valuable, it is essential that the data it produces is required and beneficial to the intended users. Effective indicators offer useful information to a broad spectrum of stakeholders.
☐	Actionable	An evaluator should make sure that the data and knowledge received from the indicator are utilized and that such information influences planning and decision-making. Since measuring an indicator can demand significant time, resources, and expense, it is crucial that the results are actionable, potentially guiding policy development, decision-making, or resource distribution.
☐	Aligned	Before deciding to measure a particular indicator, it is vital to confirm that the information it seeks is not already obtainable from another source. If similar indicators exist, it is important to coordinate with other organizations to ensure consistency in measurement practices, systems, and timelines.

Source: Adapted from UNAIDS/Monitoring and Evaluation Reference Group (MERG), pages 11–12.

Criteria 2: The Indicator Must Demonstrate Technical Merit

An indicator with technical merit ensures that the data collected is reliable, accurate, and credible. Technical merit refers to the ability of an indicator to consistently provide high-quality data under different conditions. An indicator with strong technical merit is reliable, specific, and peer-reviewed to confirm its validity.

Criteria 2: The indicator must demonstrate technical merit		
	Assessment Points	**Explanation for Each Assessment Point**
☐	Technically reliable and significant	The indicator must have technical reliability and significance, measuring something of value within its domain. It should provide clear, focused data and allow for straightforward interpretation of changes. The indicator should also be sensitive enough to detect variations in performance.
☐	Sensitive	The indicator should demonstrate monitoring merit by being reliable and sensitive, consistently producing similar results even when different instruments, procedures, or observers are used. This ensures a low margin of error.
☐	Accurate and specific	The indicator must measure exactly what it is intended to measure, with no ambiguity or overlap with other indicators.
☐	Clear	The indicator should be clear and easy to interpret, leaving no room for multiple interpretations.
☐	High quality	The indicator should undergo a thorough peer review process to ensure its quality. This review may involve experts in monitoring, evaluation, and technical fields, and can include the formation of review panels to validate the indicator's technical merit.

Source: Adapted from MERG, pages 12–13.

Criteria 3: The Indicator Must Be Clearly Defined

A well-defined indicator is critical for consistent application in the evaluation process. It ensures that everyone involved in data collection and analysis understands the indicator's purpose, how it is measured, and how to interpret its results. Without clarity, data collection may lead to inconsistency and inaccurate conclusions.

Criteria 3: The indicator must be clearly defined		
	Assessment Points	**Explanation for Each Assessment Point**
☐	Title and Definition	The title of the indicator should be concise yet descriptive, summarizing the essence of the indicator. This facilitates easy identification and referencing during daily operational and monitoring activities.
☐	Purpose and Rationale	Provide a detailed statement that explains the purpose of the indicator, including why it is necessary within the health policy context. This should highlight the importance of the indicator in tracking progress, informing decisions, or evaluating outcomes.
☐	Measurement Method	Describe the approach used to measure the indicator, including any calculations involved. Clearly specify the numerators and denominators if applicable, as well as any formulas used, to ensure consistent application and accuracy in measurement.
☐	Data Collection Method	Outline the process for gathering data needed for the indicator, including where the data will come from—for example, surveys, administrative records, electronic health records. Clarify who will be responsible for data collection and the tools or systems that will be used.
☐	Measurement Frequency	Define how frequently the indicator will be measured—weekly, monthly, quarterly, or annually. Ensure that the frequency aligns with the data collection method and the needs of the health policy for timely decision-making.
☐	Level of Disaggregation	Specify how data will be broken down to provide detailed insights—for example, by age, gender, geographic location. This disaggregation allows for targeted analysis, ensuring that the indicator can provide meaningful and actionable information.
☐	Interpretation	Offer clear guidelines on how to interpret changes in the indicator values. For example, explain what an increase or decrease in the indicator signifies and provide context for understanding these shifts. If the indicator can be interpreted in multiple ways, clarify how to differentiate between them.
☐	Strengths and Weaknesses	Identify common challenges associated with measuring this indicator, such as data quality issues, limited data availability, or potential biases. Provide practical advice or solutions to address these weaknesses to improve the indicator's reliability.

☐	Additional Sources of Information	Include relevant background information, examples of how the indicator has been applied in similar contexts, and references to related technical documentation. This section should provide stakeholders with additional resources for better understanding and application of the indicator.

Source: Adapted from MERG, pp. 13–14.

Criteria 4: The Indicator Must Be Practical to Collect and Analyze

For an indicator to be effective, it must not only be technically sound but also feasible to collect and analyze using available resources and systems. Adequate data collection infrastructure, financial and human resources, and alignment with national monitoring systems are all necessary for the indicator to be practical in real-world settings.

Criteria 4: The indicator must be practical to collect and analyze		
	Assessment Points	**Explanation for Each Assessment Point**
☐	Adequate	To effectively measure an indicator, robust systems and mechanisms must be in place, including data management platforms, reporting tools, health information systems, and well-trained personnel. These components are essential for ensuring accurate data collection, analysis, and utilization. When considering new indicators, it is important to assess whether the current infrastructure can handle these needs or if it can be adapted without significant investments. For example, integrating a new question into an existing household survey can be a cost-effective way to collect data without creating a new system. This approach leverages existing resources, reducing the need for extensive additional investments.
☐	Aligned	Indicators should be developed in alignment with existing national monitoring and evaluation systems to ensure a coherent approach. Designing indicators that harmonize with current national frameworks helps to prevent redundancy and overlap with other evaluation processes. This alignment ensures that the data collected is relevant and comparable across different systems, facilitating more efficient use of resources and avoiding conflicting or duplicative data efforts. Coordinated development of indicators also helps streamline data collection, making it more manageable and less resource-intensive.
☐	Resource availability	Sufficient financial and human resources are crucial for the effective measurement of indicators. This includes funding for data collection, training personnel, and resources for data analysis and reporting. It is essential to weigh the costs against the expected benefits of collecting the indicator to ensure that resources are used efficiently. Cost-effectiveness assessments can help to determine whether the indicator justifies the investment required. This step is particularly important to avoid the common pitfall of recommending numerous indicators without considering the financial and operational burdens they may impose on the system.

Source: Authors, adapted from MERG, pages 14–15.

Criteria 5: The Indicator Should Be Field-Tested or Used in Practice

Even well-designed indicators can face unforeseen challenges when implemented in real-world conditions. Therefore, it is essential that new indicators are thoroughly field-tested before being fully introduced. Additionally, all indicators should be reviewed regularly to ensure they remain relevant and effective as circumstances change.

Criteria 5: The indicator should be field-tested or used in practice		
	Assessment Points	**Explanation for Each Assessment Point**
☐	Field-Tested	Indicators, especially new ones, need to be field-tested to ensure their effectiveness and reliability in real-world settings. An indicator that seems robust in theory may encounter various challenges when applied in practice, such as difficulties in data collection, unforeseen external influences, or problems with measurement consistency. Field-testing helps to identify these potential issues early, allowing for adjustments before full-scale implementation. This testing phase ensures that the indicator can be practically applied, produces meaningful data, and aligns well with operational realities. For existing indicators, practical application in the field acts as an informal validation process, helping to confirm their continued relevance and reliability in changing environments.

☐	Regularly reviewed	Indicators should be regularly reviewed to ensure that they remain relevant and effective as conditions change. Regular assessments can uncover problems with data collection methods, such as inaccuracies, biases, or data gaps, as well as challenges in how data is interpreted. Circumstantial changes like the introduction of new medical treatments, shifts in program priorities, or evolving health policies can impact the utility and accuracy of indicators. These reviews allow stakeholders to make necessary adjustments, such as updating definitions, refining measurement methods, or even removing indicators that are no longer applicable. Continuous review is crucial for maintaining the quality and usefulness of the indicators in the face of dynamic healthcare environments.

Source: Authors, adapted from MERG, page 15.

Criteria 6: The Indicator Set Should Be Coherent and Balanced Overall

A coherent and balanced set of indicators is essential for providing a comprehensive understanding of the health policy being evaluated. A well-constructed indicator set should cover all critical aspects of the policy and include a range of indicators that assess inputs, outputs, outcomes, and impacts. The indicator set should not only reflect the policy's goals but also be practical and appropriate for the specific context in which it is used.

Criteria 6: The indicator set should be coherent and balanced overall		
	Assessment Points	**Explanation for Each Assessment Point**
☐	Adequate	The set of indicators should provide a thorough and comprehensive assessment of the health policy's performance. This means the indicators collectively should cover all necessary areas to give a complete picture of the policy's effectiveness, strengths, and weaknesses. An adequate set captures all key facets of the policy, including implementation processes, service delivery, and final outcomes. Without a comprehensive set of indicators, the evaluation may miss critical insights, leading to an incomplete understanding of the policy's true impact.
☐	Relevant	Relevance means that each indicator in the set should directly relate to important aspects of the policy being assessed. The indicators should address the key questions of the evaluation, such as whether the policy is achieving its intended goals, addressing priority health issues, and meeting the needs of the population. A relevant set avoids the pitfall of focusing too much on certain areas (like clinical outcomes) while neglecting others such as access to services or quality of care. For instance, if a policy aims to improve healthcare access, but most indicators only measure treatment outcomes, the evaluation may overlook whether access issues have been effectively addressed.
☐	Appropriate	Appropriateness refers to ensuring that the indicators encompass various levels of monitoring and evaluation, including inputs (resources invested), processes (how the policy is implemented), outputs (services delivered), outcomes (changes in health status), and impacts (long-term effects on public health). This multilevel approach allows for a nuanced understanding of how the policy functions at different stages. However, it is important that the number and type of indicators are practical and feasible given the available resources, including financial, human, and data capacity. An overly complex set with too many indicators can strain resources and lead to poor data quality, while a very minimal set might fail to capture the full scope of the policy's impact.

Source: Adapted from MERG, pp. 15–16.

Annex 7: A Sample Informed Consent Template for Health Policy Evaluations

Section	Details	Notes for Customization
Title of the Evaluation	[Insert Evaluation Name, for example, "Evaluation of [Policy Name] Implementation"]	Replace with specific policy or program name for instance, "Saudi Vision 2030 Health Policy Assessment."
Purpose of the Evaluation	We are conducting this evaluation to assess the effectiveness, efficiency, and impact of [Policy Name] on health outcomes, ensuring it improves public health and meets stakeholder needs globally. Your participation will provide evidence to guide health policy decisions, in accordance with international standards.	Specify policy – for instance, "Saudi National Health Strategy" – aligning with "assess" and "health outcomes."
What Participation Involves	Voluntary participation involving [describe activities, for example, "surveys, interviews, or observations"].	Tailor activities, time, and location – for example, "Arabic-language focus groups, 45 minutes, or community centers in Saudi Arabia." Ensure cultural and linguistic accessibility for diverse populations.
	Approximately [time, for example, "30–60 minutes"] at [location/method, for example, "online, in-person, or community setting"].	
	Data collection on [for example, "demographics, health status, or policy implementation experiences"], used solely for this evaluation, per international ethical guidelines.	
Confidentiality and Data Security	Protect privacy by keeping personal and sensitive information confidential, using anonymization and de-identification techniques where applicable, with access restricted to authorized personnel, National Committee of Bioethics (NCBE).	Add local or regional data protection laws, for example, "Ensure compliance with Saudi health regulations."
	Implement robust data security measures, including encryption, secure storage, and restricted access, to prevent breaches and maintain data integrity, in compliance with Saudi Personal Data Protection Law, 2021, if applicable.	
Risks and Benefits	Minimal risks, such as discomfort discussing sensitive topics, mitigated by support resources – for instance, counseling, helplines – and ethical review processes.	Specify support resources, for example, "Local counseling at [number]" or regional benefits, such as "Saudi Vision 2030 health goals."
	Potential benefits include contributing to improved global or local health policies, enhancing community well-being, and providing insights that may benefit your organization or population.	
Your Rights	Participation is entirely voluntary; you may withdraw at any time without penalty or impact on services, per the Declaration of Helsinki.	Include local ethics contacts for example, "Saudi Health Council Ethics Committee," and language options, such as "Available in Arabic, English, or local dialects." Ensure accessibility for low-literacy or disabled individuals.
	Refuse to answer any questions or stop participation at any point.	
	Receive a copy of this form, translated into your preferred language if needed, and contact [Evaluation Coordinator] at [Email/Phone] for questions or concerns. For ethical issues, contact [Ethics Committee] at [Email/Phone].	

Informed Consent	I have read and understood the information above, or it has been explained to me in [language, for example, Arabic/English], and I voluntarily agree to participate. I understand my rights, the evaluation's purpose, data use, and privacy protections, per international ethical standards. Name of Participant: _____ Signature or Mark: _____ Date: _____	Specify languages for example, "Arabic or English," or alternative consent methods such as verbal consent or thumbprint for illiterate or disabled participants.
For Participants Unable to Sign	If unable to sign but agreeing, a witness or authorized representative may sign: Witness/Representative Name: _____ Witness/Representative Signature: _____ Date: _____ Relationship to Participant: _____	Include only if targeting populations with literacy or physical barriers, such as rural Saudi Arabia.
Contact for Questions or Concerns	For evaluation questions, contact [Evaluation Coordinator Name], [Title], at [Email Address] or [Phone Number]. For ethical concerns, contact [Ethics Committee Chair Name], [Title], at [Ethics Email Address] or [Ethics Phone Number].	Include contact details for a designated evaluation lead – for example, from the SHC or Ministry of Health and an ethics contact – NCBE.

Annex 8: Consolidated Evaluator Checklist for the Six-Step Evaluation Process

Step 1: Identify and Engage Stakeholders

#	Checklist Item	✓	If No, Specify
1.1	Have all relevant stakeholders been identified?	☐	
1.2	Have stakeholders been categorized into the following main groups? • those impacted by the policy • those implementing the policy • those using the evaluation findings	☐	
1.3	Have key stakeholders who influence credibility, implementation, institutionalization, or decision-making been identified?	☐	
1.4	Have the interests and influence of stakeholders been analyzed and mapped?	☐	
1.5	Has there a stakeholder engagement plan outlining when and how to involve them?	☐	
1.6	Has each stakeholder's engagement in key policy evaluation steps—describing health policies, designing evaluation, collecting and analyzing data, justifying conclusions, using and disseminating the evaluation findings—been identified and agreed?	☐	

Step 2: Describe the Health Policy

#	Checklist Item	✓	If No, Specify
2.1	Have all key policy components—inputs, activities, outputs, outcomes, and impacts—been identified?	☐	
2.2	Are the policy components clearly described and understood?	☐	
2.3	Has a logical map been developed and visualized to show the relationships among the policy components?	☐	
2.4	Has the political, social, and financial context of the policy been analyzed?	☐	

Step 3: Design Evaluation

#	Checklist Item	✓	If No, Specify
3.1	Has the focus of evaluation been determined?	☐	
3.2	Has the type of evaluation required been identified?	☐	
3.3	Have the data needs been determined?	☐	
3.4	Have the most suitable evaluation methods been selected given the following key parameters? 3.4.1 Data needs 3.4.2 Available resources 3.4.3 Policy context 3.4.4 Evaluation comprehensiveness 3.4.5 Timing	☐	

Step 4: Data Collection and Analysis

#	Checklist Item	✓	If No, Specify
4.1	Have the evaluation questions been identified?	☐	
4.2	Have the indicators needed to measure each evaluation question been identified?	☐	
4.3	Are the indicators aligned with the national health indicators?	☐	
4.4	Have the indicators been chosen from an existing inventory or newly developed?	☐	
4.5	If newly developed, have new indicators been pilot-tested to reduce measurement error?	☐	
4.6	Is the necessary data available and are data sources appropriate?	☐	
4.7	Which data collection methods best suit the evaluation context and objectives?	☐	
4.8	Have quality and quantity issues in data collection been evaluated considering the resource constraints?	☐	

Step 5: Justify Conclusions

#	Checklist Item	✓	If No, Specify
5.1	Has the policy context been considered in data interpretation?	☐	
5.2	Are findings compared with literature and similar policies?	☐	
5.3	Have alternative explanations been explored and tested?	☐	
5.4	Are benchmarks – Vision 2030 or transformation goals – used to assess results?	☐	
5.5	Are results compared with previous years' outcomes?	☐	
5.6	Are actual findings compared to intended policy goals?	☐	
5.7	Have potential biases (if any) in the analysis been documented?	☐	
5.8	Have the limitations (if any) of the evaluation been fully assessed?	☐	

Step 6: Use and Disseminate Evaluation Findings

#	Checklist Item	✓	If No, Specify
6.1	Is there a strategy to promote use of the evaluation findings?	☐	
6.2	Are evaluation findings used to inform both short-term and strategic planning?	☐	
6.3	Are there regular meetings to communicate the evaluation conclusions?	☐	
6.4	Are evaluation reports tailored to suit different stakeholder audiences?	☐	
6.5	Are findings communicated clearly and in a timely manner?	☐	
6.6	Are findings disseminated through various accessible methods?	☐	


WORLD BANK GROUP
المجلس الصحي السعودي
Saudi Health Council