The Best Practice Handbook

Diagnosing Root Causes and Driving Results

Office of the Legislative Auditor General

Report to the **UTAH LEGISLATURE AND UTAH GOVERNMENT ORGANIZATIONS**

LEGISLATIVE AUDITOR GENERAL

LEGISLATIVE
AUDITOR GENERAL

GOVERNMENT
EXCELLENCE

Audit Subcommittee

President J. Stuart Adams, Co-Chair
President of the Senate

Speaker Mike Schultz, Co-Chair
Speaker of the House

Senator Kirk A. Cullimore
Senate Majority Leader

Representative Casey Snider
House Majority Leader

Senator Luz Escamilla
Senate Minority Leader

Representative Angela Romero
House Minority Leader

Audit Staff

Kade R. Minchey, Auditor General, CIA,
CFE

Jesse Martinson, Manager, CIA

Brent Packer, Audit Supervisor

Dr. David Yokum, Ph.D, Consultant

Audit Subcommittee of the Legislative Management Committee
President J. Stuart Adams, Co-Chair | Speaker Mike Schultz, Co-Chair
Senator Kirk A. Cullimore | Representative Casey Snider
Senator Luz Escamilla | Representative Angela Romero

January 30, 2026

TO: THE UTAH STATE LEGISLATURE

Transmitted herewith is our report:

"The Best Practice Handbook: Diagnosing Root Causes and Driving Results"
[Report #2026-04].

The Best Practice Handbook: Diagnosing Root Causes and Driving Results is a resource for all government organizations in the state of Utah. In the Handbook, we describe the Five Layers of Cause framework for diagnosing the root causes of organizational issues. *Governance, Leadership, Culture, Structure*, and *Process* are the five layers that influence outcomes in every organization. The Handbook can assist leaders to accurately identify which causes are driving poor performance. Leaders can use the knowledge gained to intervene on the most critical causes of organizational issues. This approach drives results by providing effective and sustainable solutions.

We believe cause is the key to effective and efficient government. It tells leaders where and how to intervene. The result is better outcomes, lower costs, enhanced decisions, and healthier organizations. We encourage all leaders to apply the principles and practices from the Handbook. Doing so will help to maximize the value of government products and services for the benefit of Utah citizens.

This report is published initially under an "Open Comment Period." The purpose of the open comment period is to receive feedback from a broad audience on how to make this report as useful and practical as possible. Once the open comment period concludes, we will publish a final version based on feedback we have received.

Sincerely,

Kade Minchey

Kade R. Minchey, CIA, CFE

Auditor General

kminchey@le.utah.gov

Table of Contents

BEST PRACTICE HANDBOOK

▶ BACKGROUND

This is the third Best Practice Handbook released by our office. *The Best Practice Handbook: A Practical Guide to Excellence For Utah Government* and *The Best Practice Handbook For Student Achievement in Public Education* have helped many government organizations identify and mitigate potential pitfalls.

This Handbook focuses on diagnosing the root causes of organizational issues. As a doctor relies on their diagnosis of disease to prescribe an effective treatment, organizational leaders must use the knowledge gained from root cause analysis to intervene effectively on issues.

The goal of every government leader is to maximize the value of products and services for the benefit of citizens. The Five Layers of Cause framework assists leaders in this goal. The results are better outcomes, lower costs, enhanced decisions, greater impact, and healthier organizations. We encourage leaders to apply these methods to improve their organizations.

THE BEST PRACTICE HANDBOOK: DIAGNOSING ROOT CAUSES AND DRIVING RESULTS

THE FIVE LAYERS OF CAUSE FRAMEWORK

🏛 Governance
👤 Leadership
👥 Culture
🏢 Structure
⚙ Process
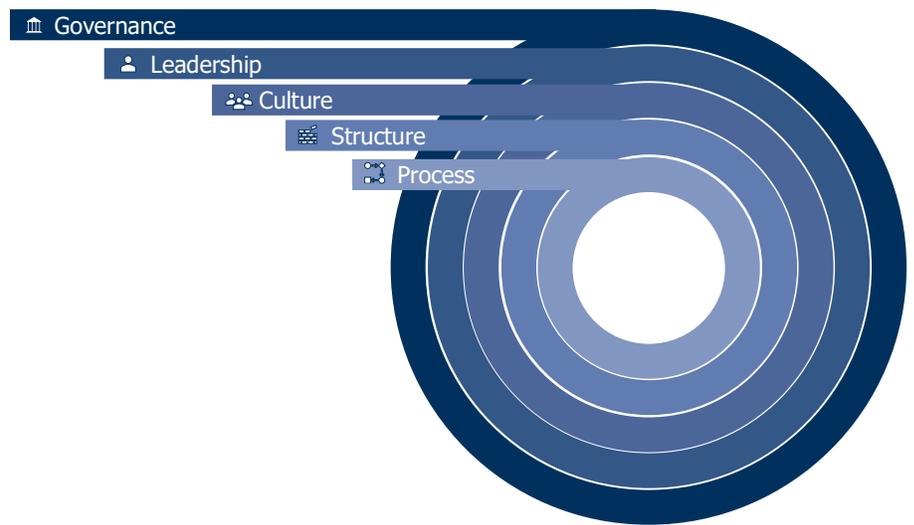
⌐ THE FIVE LAYERS OF CAUSE

We developed The Five Layers of Cause framework to assist leaders in identifying the causes of organizational success and failure.

- **Governance —** The laws and rules that establish the organization's mission, accountability, and funding
- **Leadership —** The competency and character of leadership
- **Culture —** The collective beliefs, values, and behaviors of the organization
- **Structure —** The policies and standards that guide work
- **Process —** The steps taken to accomplish a task

Leaders should use the knowledge gained from the framework to intervene effectively on organizational issues.

Section 1
Accurate Diagnosis of Organizational Issues Is Essential to Promote Organizational Health

In 2023, our office released *The Best Practice Handbook: A Practical Guide to Excellence for Utah Government* (BPH).[1] The principles from the BPH have helped many government organizations proactively identify and mitigate potential pitfalls. It further increased the state's use of best practices.

This report, *The Best Practice Handbook: Diagnosing Root Causes and Driving Results*, furthers the goals of the BPH by focusing on how to evaluate organizational health. The BPH promotes the use of a continuous improvement framework known as The Management Cycle. Through continual planning, implementing, and evaluating, organizations can systematically improve their products and services. This Handbook focuses exclusively on the third phase of The Management Cycle—Evaluation. This Handbook digs deeper into *how* to effectively evaluate government policies, programs, products, and services. Because evaluation is the driving force for planning and subsequently implementing, it is a critical leverage point in organizational improvement. We believe that dedicating resources and building capacity to increase the frequency and quality of evaluations will provide a powerful return on investment, ultimately maximizing the value of the taxpayer's dollar.

Stakeholders Need to Know What Works in Government

To truly improve as an organization, leaders must understand the causes of major risks. Only when leaders thoroughly understand these causes can they make meaningful improvements. We spoke with key policy- and decision- makers throughout Utah government to learn what types of information are most valuable to them. Stakeholders consistently need information on how to make government work well for the citizens of the state. They want to know the causes of issues so that they can intervene effectively.

---

[1] *The Best Practice Handbook: A Practical Guide to Excellence for Utah Government* (Report No. 2023–05), Office of the Legislative Auditor General.

They also want to know if policies, programs, and initiatives are successful. If successful, stakeholders want to know how to expand. If unsuccessful, they want to make changes to promote better outcomes. In short, stakeholders are interested in the type of information that enables them to make government better.

One stakeholder we spoke to told us they are drowning in data and performance measures, yet frequently they do not have the information they need to make good decisions. We believe there has been too much focus on collecting data and not enough focus on using that data to provide meaningful information to stakeholders. Good evaluation bridges this gap by translating data points into actionable decisions.

> **One stakeholder told us they are drowning in data and performance measures but lack the information needed to make good decisions.**

While the subject of information needed varies by organization, the type of information does not. Causal information is generally most valuable to stakeholders.

Causal Information Is Essential

Causal information moves beyond describing the state of the world to knowing how to change it.

Understanding cause allows us to maximize the impact of our efforts. Causal information

- Tells us how to improve outputs and outcomes
- Enables us to avoid waste and optimize efficiency
- Supports better forecasts and predictions
- Helps to mitigate risks
- Gives policy- and decision-makers answers to their core questions

Organizations must simultaneously deliver products and services and evaluate the efficiency and effectiveness of those products and services. Dr. W. Edwards Deming, a prominent organizational theorist, taught, "Every system is perfectly designed to get the result that it does." Determining *why* processes, programs, initiatives, and interventions are getting the results that they are should be the first step in any improvement effort.

There are two main approaches to determining the causes of results: root cause analysis (RCA) and causal reasoning. RCA identifies the symptoms of problems then traces backward to determine what led to the negative outcome. The second approach, causal reasoning, designs a solution then carefully measures the impact the change had. Figure

1.1 illustrates the two approaches used to find the cause and determine what is and isn't working well in government.

Figure 1.1. Two Main Approaches Are Used to Determine What Works in Government. Root cause analysis identifies symptoms of poor organizational health, then traces backward to determine what conditions led to the outcome. Causal reasoning designs solutions and carefully measures the impact these interventions have.

1. Root Cause Analysis   ## 2. Causal Reasoning

Source: Auditor created.

Both needs—understanding the causes of issues and the effects of interventions—are critical to success.

Section 2 of this report focuses on the first need, RCA. Section 3 focuses on the second need, causal reasoning. The remainder of this Section lays the foundation for thinking about cause and effect.

> **Good strategy requires an understanding of the causes of issues and the effects of interventions.**

Root Cause Analysis Drives Greater Value from Taxpayer Investments

RCA determines why we get the results we do. The value of this information is clear—it empowers decision-makers to drive better outcomes at lower costs. Evaluations should aim to produce this type of information wherever possible.

RCA offers a high return on investment. Simply put, RCA helps answer questions that begin with "*Why*." For example, a government leader might ask the following questions:

- *Why* did underage drinking go up the past year?

- *Why* is recidivism increasing?

- *Why* do we have a backlog in cases?

- *Why* are student learning outcomes lagging in our school?

To answer these questions, we must determine *the causes* of the current results. More specifically, we must determine the *root causes* of results. Root causes are deeper, more fundamental causes of organizational issues. Answers to these questions enable us to craft effective interventions and make accurate predictions.

Causal Reasoning Demonstrates What Works in Government

The second approach, causal reasoning, looks at an intervention and determines its impact. Often this involves a *design* element. This means evaluators should determine how success will be measured prior to implementation. In some circumstances, statistical techniques can be used after implementation to reliably determine the impact of an intervention. Causal reasoning can answer questions such as:

- Did the policy change have the intended effect?
- Should the pilot program be expanded?
- Is our strategy effective?
- Where can money best be allocated?

This approach provides information on the *impact* of policies, programs, and other interventions. Causal reasoning

- Enables stronger predictions
- Supports effective interventions
- Promotes greater efficiency through elimination of unproductive activities

When paired together, RCA and causal reasoning can help leaders make good decisions. They further offer extraordinary opportunities to solve problems and improve strategies.

What Is Cause and Effect?

Effects have causes. In the medical field, doctors rely on an understanding of cause and effect to diagnose and prescribe the right treatments for their patients.
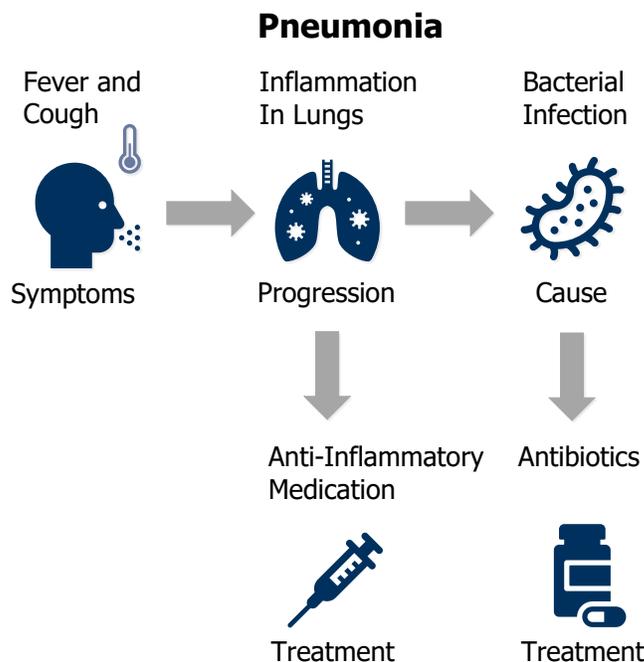
Cause and effect are used to diagnose and prescribe effective interventions.

Figure 1.2 illustrates how this approach is used in medicine.

Figure 1.2. Doctors Use Clues from a Patient's Symptoms to Diagnose and Treat the Causes of an Illness. Similarly, leaders can use organizational symptoms to understand the causes of good and poor performance.

Pneumonia

Fever and Cough	Inflammation In Lungs	Bacterial Infection
Symptoms	Progression	Cause

Anti-Inflammatory Medication

Antibiotics

Treatment

Treatment

Source: Auditor created from "National Institutes of Health: Pneumonia Causes and Risk Factors" and "Mayo Clinic: Diseases and Conditions (Pneumonia)."

As medical professionals have improved their ability to determine the causes of symptoms, they have been able to intervene more effectively. For example, pneumonia caused by a bacterial infection is often treated with antibiotics. Antibiotics stop the bacteria from multiplying, ultimately preventing serious illness and promoting a speedy recovery. In contrast, ignoring cause and giving the patient a cough drop and pain reliever may alleviate discomfort, but it will not resolve the source of the problem.

Similarly, various symptoms can point to underlying issues in organizations. Our office studied these relationships between organizational issues and their causes. We found
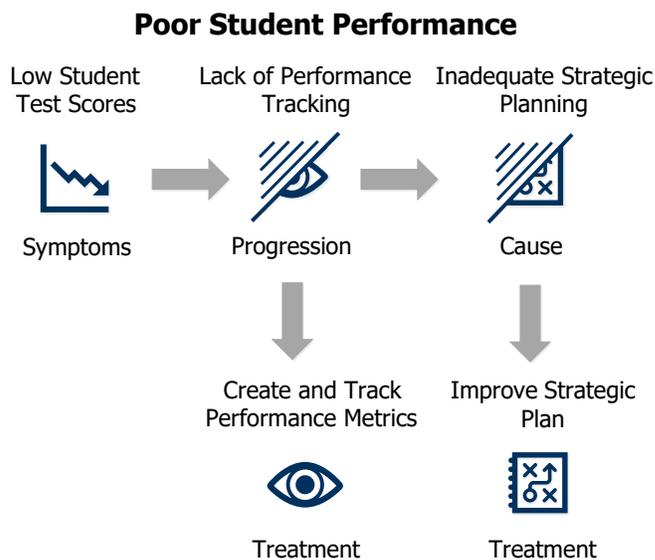
that symptoms and causes can be mapped and referred to in solving organizational issues.[2,3] The next subsection compares the approach used by doctors to organizations.

Symptoms Play a Crucial Role in Determining Root Causes

Symptoms are the effects of negative conditions. When an organization exhibits symptoms, it offers clues as to what might be causing those symptoms. For example, the Utah Schools for the Deaf and the Blind (USDB) students consistently achieved under 15 percent proficiency on standardized assessments from 2019 to 2024.[4] However, this low performance could be traced back to insufficient performance tracking and inadequate strategic planning as possible causes. These factors acted as contributors in producing a negative outcome. Figure 1.3 illustrates how negative conditions lead to poor outcomes.

Note: Examples are marked by a blue vertical line on the lefthand margin.

Figure 1.3. Poor Student Performance Is Preceded by Negative Conditions. These conditions contribute to causing the adverse result. By determining the underlying causes of an issue, interventions can be crafted to effectively reach a resolution.

Poor Student Performance

Low Student Test Scores	Lack of Performance Tracking	Inadequate Strategic Planning
Symptoms	Progression	Cause

	Create and Track Performance Metrics	Improve Strategic Plan
	Treatment	Treatment

Source: Auditor created.

---

[2] Though we primarily discuss cause and effect in terms of diagnosing illness in this report, they can also be applied to organizational wellness. When taken this way, organizational leaders can determine the factors that led to success. This informs how to replicate and expand efforts to improve other policies, programs, and initiatives.

[3] As in the medical field, diagnosis of organizational issues is part science, part art. Consequently, our framework is a guide rather than a math equation. This is why it is called *the practice* of medicine in the medical field and *the practice* of root cause analysis here.

[4] Based on RISE and Utah Aspire Plus assessments. Excludes 2020 for which proficiency data is not available.
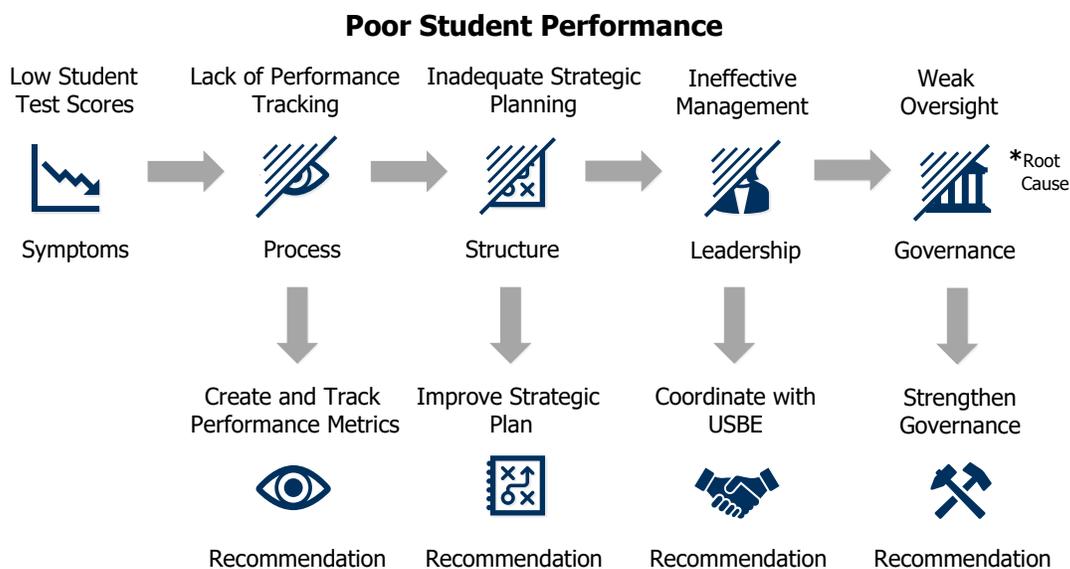
Our 2025 audit[5] found that USDB did not have an organization-wide strategic plan. Strategic plans for the School for the Deaf and the School for the Blind were also lacking several core components. Because each school had a separate plan, it created a risk of mission misalignment at USDB. These issues in planning contributed to poor performance tracking and evaluation, and ultimately poor student performance.

Poor performance in any environment is the result of negative conditions. Our role as leaders, auditors, and evaluators is to understand what those conditions are and to intervene to address the underlying causes. When we focus on cause, we can prevent serious issues from materializing. We can also promote efficient improvement. In contrast, when we intervene at the symptom level, the problem may persist for long periods of time and even worsen.

We call this relationship between symptoms and causes "symptom-cause pairs." In the example just described, only two causes are connected to a single effect. However, symptom-cause pairs often include several causes. For example, pneumonia might also be caused by a virus or even fungi. Similarly, poor strategic planning may be one of several causes driving poor performance, as seen in Figure 1.4.

Figure 1.4. Poor Student Performance Is Preceded by Multiple Underlying Causes. By pinpointing the underlying causes of an issue, interventions can be designed to address the most important causes.[6]

Poor Student Performance

Low Student Test Scores	Lack of Performance Tracking	Inadequate Strategic Planning	Ineffective Management	Weak Oversight
Symptoms	Process	Structure	Leadership	Governance *Root Cause
	Create and Track Performance Metrics	Improve Strategic Plan	Coordinate with USBE	Strengthen Governance
	Recommendation	Recommendation	Recommendation	Recommendation

Source: Auditor created.

[5] *A Performance Audit of the Utah Schools for the Deaf and the Blind* (Report No.2025–22), Office of the Legislative Auditor General.

[6] The audit report notes that USDB students may not achieve the same level of proficiency as their peers without a disability. This is not included in the graphic as it lies outside of leadership's control.

As seen in Figure 1.4, a governance issue was the root cause leading to poor student performance. It is essential to identify and intervene on the root cause of an issue because if we do not, we will not fix or cure the problem. At the same time, if we recommend only fixing the root cause, we won't help the organization as much as we could. To improve the organization or heal it, we need to look at all causes driving an issue. Like a doctor that prescribes antibiotics for pneumonia to eliminate a bacterial infection (the root cause), leaders should intervene in the governance and leadership causes of organizational issues wherever possible. At the same time, as a doctor prescribes anti-inflammatory medication to reduce pneumonia's progression (surface cause), leaders should also intervene on other identified causes to limit organizational issues from becoming pervasive.

> **Problem symptoms often share common causes. Recognizing symptom-cause pairs can speed up and improve root cause analysis.**

As previously noted, USDB struggled with performance tracking and strategic planning. The USDB audit report also noted poor financial management for more than a decade. Combined, these concerns create a fuller picture of the performance issue and offer high-impact opportunities to intervene in the deeper causes undermining outcomes. In Section 2, we will provide greater detail on how to identify root causes by introducing The Five Layers of Cause framework. While addressing every cause may not be feasible, understanding the set of factors that influence results will provide greater flexibility and opportunity to leaders.

Many of the challenges organizations face can be viewed as symptoms of deeper causes. For example,

- Process delays could be the result of a disengaged culture.
- Inadequate staffing levels could be a sign of toxic culture.
- Poorly designed policies could suggest incompetent leadership.
- Widespread noncompliance might signal inadequate governance.

By identifying and addressing deeper layers of cause, the organization can effectively eliminate issues. When we understand the causes for the results we are getting, we can develop systems that produce the outcomes we want.

> **When we understand the causes for the results we are getting, we can develop systems that produce the outcomes we want.**

As organizations identify the factors that drive or deter success, they can meaningfully advance their missions. Government leaders can use this information to maximize the value of their government products and services. We hope that by providing this Handbook, we empower all

employees to improve their respective organizations and to achieve better outcomes. We now move to Section 2 where we will explore the Five Layers of Cause framework. This will help determine the depth of each issue. In Section 3, we will discuss how to effectively intervene on the identified causes of issues.

Section 2
Understanding Cause Creates the Power to Improve Government

If leaders fail to understand the causes of success and failure, they will be ineffective. The goal of every government leader is to maximize the value of products and services for the benefit of the citizens of the state. Without understanding cause, leaders are powerless to intervene on the sources of problems because they do not know what those are. They are also limited in their ability to optimize taxpayer investment because they do not know which ingredients in the causal recipe are most critical to success. Simply put, understanding cause creates the power to significantly improve government.

We developed The Five Layers of Cause framework (the framework) to assist leaders in identifying the causes of organizational success and failure. When organizations do not meet objectives or achieve goals, the framework directs stakeholders to determine why. When organizations improve outcomes and advance missions, the framework identifies what drives the difference. This empowers leaders to meaningfully advance the mission of their organizations.

An organization exists to reliably produce a set of outcomes for its stakeholders. So, leaders should specify which outcomes they aim to impact. The achievement of outcomes are the primary criteria organizations are measured against.

An organization exists to reliably produce a set of outcomes for its stakeholders.

At their simplest, organizations are a collection of people, rules, and capital. These are spread across the following five layers of cause:

- **Governance**—The laws and rules that establish the organization's mission, accountability, and funding

- **Leadership**—The competency and character of leadership

- **Culture**—The collective beliefs, values, and behaviors of the organization

- **Structure**—The policies and standards that guide work

- **Process**—The steps taken to accomplish a task

The list is hierarchical, ranking the layers from most to least impactful.

Figure 2.1 illustrates the hierarchy.

Figure 2.1. The Five Layers of Cause Framework, Ordered from Most to Least Impactful.
Good governance is vital to organizational success, while process is more limited in its impact.

🏛 Governance
👤 Leadership
👥 Culture
🧱 Structure
🔀 Process
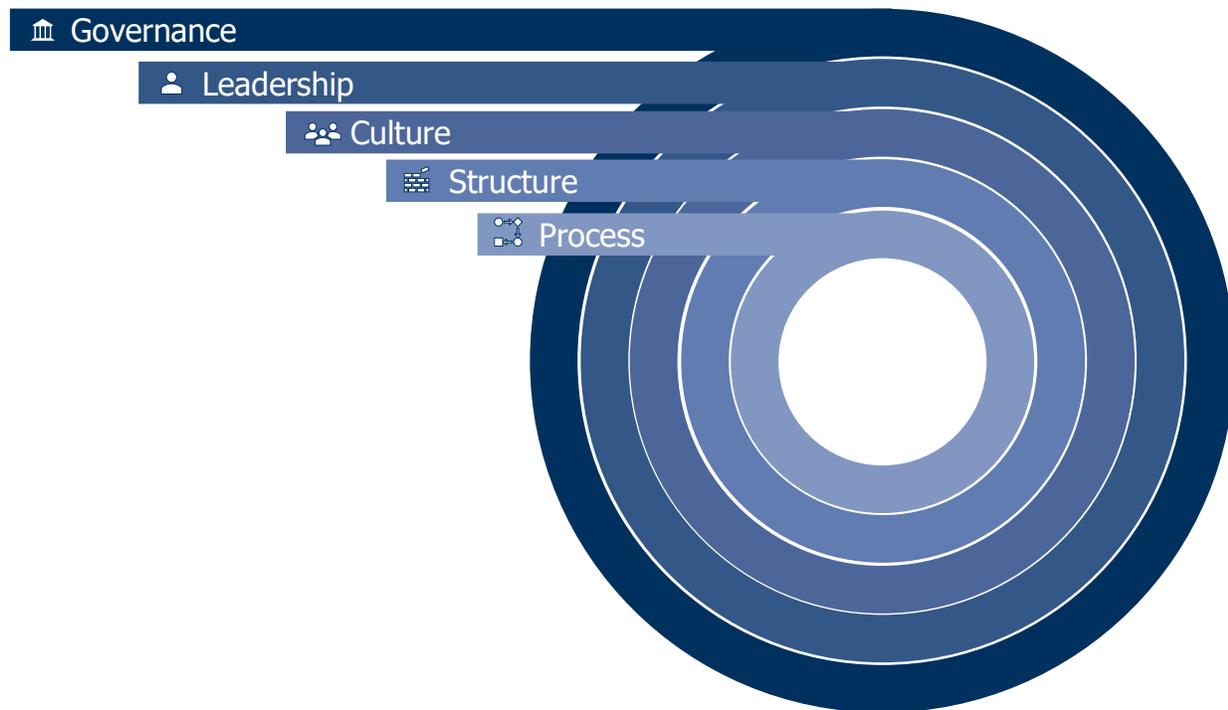
Source: OLAG's Five Layers of Cause Framework.

We believe organizations are likely to achieve outcomes with well-designed governance, effective leadership, healthy culture, and well-calibrated structures and processes. The inverse leads to unfulfilled or even negative outcomes.

When issues occur, stakeholders must accurately diagnose the problem. Generally, when an organization has a governance weakness, that weakness flows downstream to issues across all other layers. In contrast, when an issue resides at the process level, it tends to be less pervasive.

The term "effectiveness" refers to success in advancing outcomes. Leaders should assess each of the five layers of cause for how well it moves intended outcomes forward. A poorly designed or executed layer of cause undermines the organization's success.

The term "efficient" refers to getting the best return on investment. This is accomplished through eliminating waste. An organization can be highly effective yet inefficient. The opposite is also true. To maximize value, leaders must be efficient in how they advance outcomes.

Waste runs counter to every reputable organization's mission. A primary source of waste is misdirected effort. Causal analysis and causal reasoning overcome this risk by grounding operations in a correct understanding of what works. It is the method for determining how to reliably produce a set of outcomes at the lowest cost for stakeholders. The next five subsections cover how to diagnose cause at each layer of the framework.
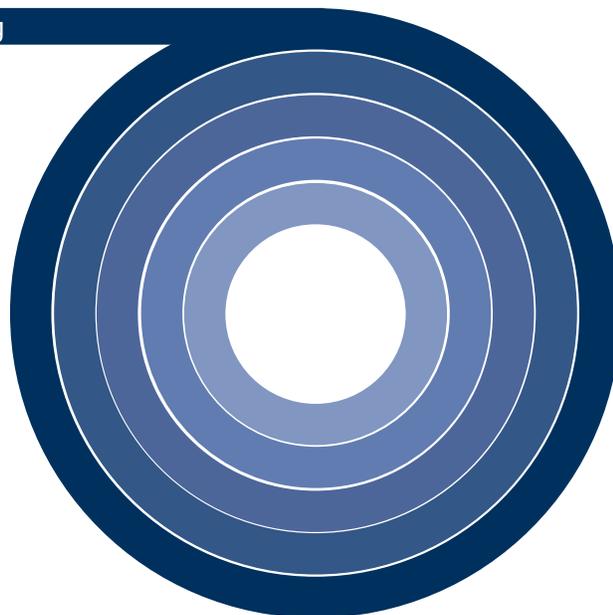
A primary source of waste is misdirected effort.

Subsection 2.1
Diagnosing Governance Issues

Governance | Mission | Accountability | Funding

🏛 **Diagnostic Checklist**

Governance – Sets the organization's mission, drives accountability, and empowers the organization with the leadership and resources needed to achieve meaningful results.

Common Deficiency	Evidence to Collect	Indicators / Metrics
☐ Mismatch between resources and priorities	Budget allocations, spending reviews	% variance between budget and mission-critical programs
☐ Weak oversight	Audit committee reports, board charters, board agendas	# of unresolved audit issues
☐ Mission drift	Statutory provisions, grant contracts, strategic plans	% mismatch between mission, funding sources, and activities
☐ Misaligned incentives	Compensation policies, performance metrics	# incentives not linked to outcomes
☐ Obsolete mandates	Statutes, rule reviews	# outdated statutes and rules
☐ Weak public or stakeholder feedback loops	Public consultations, complaint data	Frequency of feedback by type
☐ Overlapping authorities	Inter-agency charters	# duplicated oversight areas
☐ Poor disclosure practices	Annual reports, open data portals	Compliance with transparency standards

The first layer of cause is governance. If governance is broken, the organization is broken. In our search to diagnose sickness (deficiency) in an organization, we must start with governance. Good governance has the authority and power to heal (improve) an organization.

A leader, auditor, or evaluator looking to improve an organization should begin by examining governance. Similar to medical practice, diagnosis often begins by ruling out the most serious concerns, such as governance issues. One doctor we spoke to told us he treats what is the most likely cause while testing for what is most serious. If governance is an issue, all other layers of cause are likely impacted. It is essential to know if governance weaknesses are driving challenges across the organization.

Mission Drift Promotes Waste

Within governance, the first thing to evaluate is the mission. If an organization strays from its mission, then it will not reliably produce the set of outcomes it was created to produce. For example, the Utah Office of the Inspector General of Medicaid Services (OIG) is statutorily authorized to provide Medicaid oversight. Despite this provision, the OIG had provided little evidence of independent oversight of Accountable Care Organizations (ACOs) since 2018. ACOs manage nearly one-and-a-half billion dollars in Medicaid spending. By failing to exercise oversight of such a large risk, the OIG drifted from its core mission of oversight and put over one billion dollars at risk of waste.

To know if mission drift is an issue in the organization, review *Utah Code* and, where applicable, the *Utah Constitution*. Questions for assessing mission drift include the following:

- Are we serving our customers' primary needs?
- Are we advancing the core outcomes of our mission?
- Are we minimizing waste to maximize efficiency?

By anchoring the organization to the mission and to the primary needs of customers, organizations can avoid issues of mission drift.

Weak Accountability Hurts Value Creation

The second thing to evaluate in governance is accountability. Public organizations generally lack the profit motives and price signals prevalent in the private sector. Accountability ensures the value of products and services does not suffer because of this difference. Leaders must hold their people accountable for results.

To manage results, leaders must measure results. Performance management should include metrics like effectiveness and efficiency. In Section 3, we present strategies for assessing and measuring impact. Leaders should use all these measures to support decision-making. Leaders should also align incentives to promote performance excellence.

When organizations do not focus on accountability in performance, performance deteriorates. For example, one audit found that the Office of the Utah Attorney General (OAG) tracked over 416 measures. However, these measures were primarily focused on output and workload, such as the number of cases opened. Division funds also did not connect to outcomes, which exacerbated the organization's lack of focus on results. The result was a lot of activity without an understanding of the impact of these efforts.

Tip: Review past agendas for focus on outcomes.

Some things to review for assessing accountability include the following:

- Past meeting agendas for a focus on outcomes
- Performance measures for effectiveness and efficiency
- Voting records for rubber stamping

If the governing body approves all or nearly all proposals brought to it, this could signal a lack of involvement and oversight.

Having clear reporting lines is another important aspect of accountability. Establishing decision authority and the proper guardrails empowers leaders to act within their scope. Instances of duplication and fragmentation could signal a lack of clarity in the reporting structure.

Recently, an audit of Utah's System of Higher Education (USHE) showed governance struggles, which undermined the ability of USHE to work as a system. During the course of the audit, the Board adopted a new strategic plan focused on goals. The Utah Legislature used targeted funding. Within a two-year period, graduation rates within USHE institutions rose by three percentage points. This is a powerful example of governance as a contributing factor to results.

Funding Misalignment Creates Adverse Incentives

Finally, the third area of governance to review is whether funding is aligned with the mission and goals of the organization. When funding is misaligned, incentives encourage less productive behavior. For example, an audit found that over 70 percent of the Department of Health's budget was made up of federal funds. This created specific governance and accountability challenges, such as long-term sustainability concerns. Leadership allowed the funding source to dictate organizational priorities, resulting in misaligned incentives.

Tip: Review funding sources for alignment with mission-critical activities.

On the other hand, at the time of the audit, Utah Retirement Systems (URS) stated that its objective and fiduciary responsibility was to meet the assumed investment rate of return of seven-and-a-half percent while minimizing risk in the portfolio. Recognizing that its goal was not to be a "top performer" in investment returns, URS designed portfolios to emphasize downside protection. As a result, its portfolio volatility ranked as one of the lowest among its peer group.

Leaders should proactively review governance structures to ensure the organization is fulfilling its core mission, exercising oversight appropriately, and adequately managing its sources of funding. This will help to avoid many downstream issues that occur when one or more of the parts of good governance are in poor health.

Leadership | Competency | Character
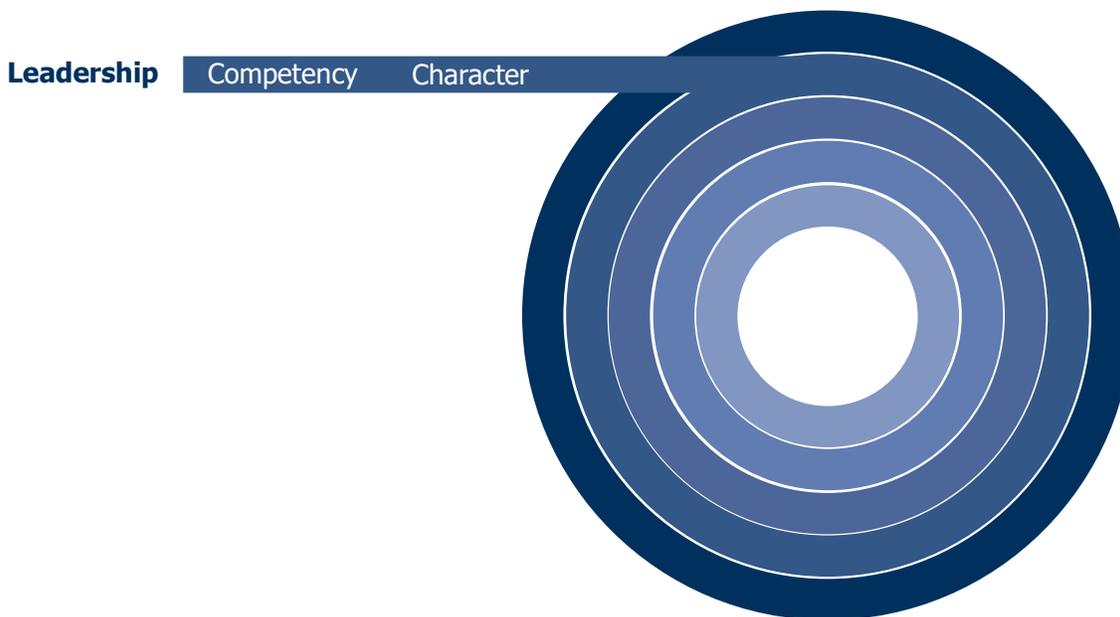
Diagnostic Checklist

Leadership – The decisive force that transforms vision into action and effort into results.

Common Deficiency	Evidence to Collect	Indicators / Metrics
☐ Lack of clear goals or objectives	Strategic plans, Workplans	# of missing components, Out-of-date workplans
☐ Weak oversight	Performance dashboards	# of outcome measures
☐ Reactive or unclear decisions	Meeting minutes, Interviews	Key decisions not documented
☐ Poor communication	Survey, Formal communications	Low survey scores, Lack of reporting
☐ Silos between units or other departments	Logic models, Memorandum of Understandings (MOUs)	Performance measures not classified, Missing MOUs
☐ Lack of evaluation	Internal Studies, Evaluations	Lack of causal reasoning tools
☐ Negative culture	Human Resource records, Incident reports, Survey	% of HR complaints by leader, % incident by unit, Low survey scores
☐ Unsatisfactory tone at the top	Ethics complaints, Survey	Serious concerns, Low survey scores

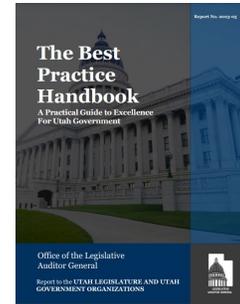Leadership is the decisive force that transforms vision into action and effort into results. We began our approach to diagnosis with governance because this is the highest authority of an organization. Ultimately, the governing body has the power to set the organization's direction. It also details how leaders are hired and terminated. But leadership is also incredibly important in the success and failure of organizations. Every

leader, auditor, or evaluator looking to assess an organization needs to understand the role of effective leadership.

The purpose of leadership is to focus, align, and mobilize organizational efforts to achieve meaningful outcomes. Effective leaders must possess essential competencies and be of good character. They are accountable for organizational performance—whether good or poor. Governing bodies should use organizational effectiveness and efficiency measures to assess the performance of leaders.

As discussed in *The Best Practice Handbook*, effective leaders skillfully apply all three phases of The Management Cycle—plan, implement, and evaluate. Leaders establish and monitor the rules, policies, and standards that guide the work. They also largely influence the culture of the organization through the tone they set at the top. This is why deficiencies in the next three layers—culture, structure, and process—often signal one or more leadership shortcomings. It is true that no organization can operate smoothly all the time. However, when deficiencies are pervasive or persistent, leaders are responsible for allowing the problem to grow and continue.

> **Effective leaders skillfully apply all three phases of The Management Cycle—Plan, Implement, and Evaluate.**

Leaders Lacking Core Competencies Undercut Organizational Health

When leaders lack core competencies, they cannot be effective. For example, an audit identified inaccuracies in the Division of Water Resources water use data. The audit found that clear inconsistencies in reported water usage indicated something was wrong with the data. However, the division continued to use the data in its municipal and industrial studies. In one case, American Fork City's water usage between 2005 and 2010 showed a 48 percent decline when, in fact, other data suggested that water usage increased. Because leadership did not demonstrate competency in evaluation, the subsequent plans and conclusions were misguided.

The three phases of The Management Cycle contain 12 best practices. These include skills such as evaluating the condition of the organization, exercising oversight, and establishing standards. Leaders should take the best practice self-assessment to determine organizational strengths and weaknesses, then systematically work to address any deficiencies. Leaders who proactively improve the organization are like patients who regularly exercise, get good sleep, and nourish their bodies to promote good health.

> **Take the Best Practice Self-Assessment.**

Though not exhaustive, some of the things we consider when assessing leadership include the following:

1. Do leaders clearly communicate direction and align people, resources, and actions to achieve measurable results?

2. Do leaders create a culture of trust, accountability, and continuous improvement where people are motivated to perform at their best?

3. Do leaders use data, feedback, and problem-solving to improve performance rather than simply react to issues?

4. Is monitoring used to identify instances of poor performance or quality?

5. When issues are identified, do leaders take action to resolve the issues?

6. Do the stated and unstated incentives encourage the desired performance?

In short, leadership is responsible for maximizing the value of the organization's products and services. When leaders do not competently apply the three phases of The Management Cycle, we believe it leads to performance declines. This undercuts the achievement of outcomes for citizens. In contrast, effective leadership can have a powerful impact on the success of an organization. Section 3 provides a compelling case study of effective leadership.

The Character of Leadership Influences the Culture of the Organization, Which Influences Outcomes

Leaders set the tone at the top of the organization. Their decisions and behaviors influence the organization's culture. Organizations operate within complex, changing environments. So, it is essential that leaders be innovative and guide their organizations to adapt.

Effective leaders assess their performance through clear, outcome-driven indicators, including the following:

- **Strategic Plan Execution**—Are priorities being accomplished or only discussed?

- **Outcome Performance Metrics**—Are results improving in measurable ways?

- **Culture & Engagement Data**—Do people feel trusted, motivated, and aligned with purpose?

- **Incident and Risk Analysis**—Where do failures cluster by leader, function, or layer of the organization?

- **Performance Management Results**—Are people growing, improving, and being held accountable?

Leaders must transform vision into action and efforts into results. Leaders should be assessed on their ability to lead the organization toward its mission. This requires the ability to focus, align, and mobilize the organization to advance outcomes. Strong leadership improves the quality of government products and services delivered to citizens.

Leaders should be assessed on how well they lead the organization toward its mission.

Subsection 2.3
Diagnosing Culture Issues

Culture   Beliefs   Values   Behaviors
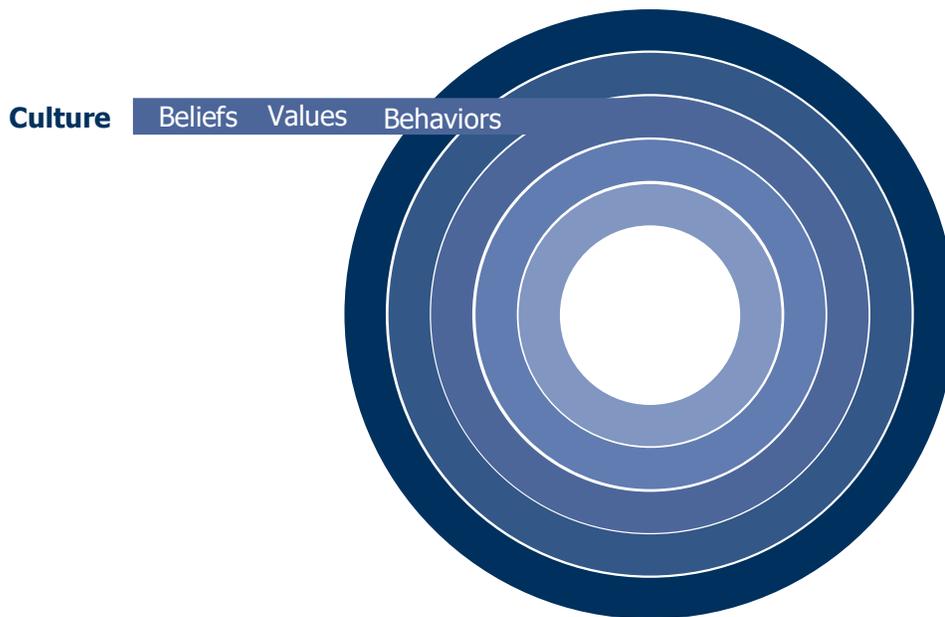
Diagnostic Checklist

Culture – The collective beliefs, values, and behaviors of the organization.

Common Deficiency	Evidence to Collect	Indicators / Metrics
☐ Noncompliance	Audit reports, Compliance logs	% of recommendations implemented, % of repeat deficiencies
☐ Low accountability	Performance appraisals, Corrective action plans	Low variation between employee scores
☐ Risk averse	Survey, Interviews	Themes of lack of innovation or punitive
☐ Low skill	Training PPTs, Employee training hours, Onboarding documents	Lack of substantive training, Low training hours, Insufficient onboarding
☐ Low productivity	Productivity ratios, Interviews	Cost-efficiency measures by team
☐ Misaligned incentives	Performance appraisals, promotions and bonus records	Mismatch between individual performance and compensation
☐ Absenteeism	Leave records	% of unscheduled vs. scheduled leave
☐ High turnover	Turnover rates	% of turnover trends

Bad culture can destroy good organizations. Leadership and culture are often connected, but they are separate diagnoses because one can be positive and the other negative.[7] The ways of addressing and diagnosing each are also different.

---

[7] For an example of this, see *A Performance Audit of the Office of the Attorney General* and *A Performance Audit of the Position of the Attorney General*. These reports found that despite a lack of involvement of the executive leader, the overall culture within the office of the attorney general was positive.

Culture Influences Structures, Processes, and the Outcomes Achieved

Culture is hard for auditors and others to evaluate, so it is often overlooked. However, one of the most powerful ways to improve outcomes is to strengthen culture. So, it is necessary to find ways to assess and improve culture. Just as leaders, auditors, and others need to be well versed in governance and leadership, they also need a deep understanding of culture. They need to understand the driving forces behind negative and positive culture. They need to know how to detect and document culture and then provide recommendations for improvement.

Leaders shape culture, but culture drives outcomes. An auditor won't find an organization's culture reviewing documents at a desk. Auditors need to be present, aware, and immersed in an organization. Culture is made up of the collective beliefs, values, and behaviors of the individuals in the organization. It is also manifest in the individual decisions, attitudes, and accountability, and the organization's structures, processes, and outcomes.

Often when performance declines, process and structure are not the root causes, although those issues may need addressing. Instead, culture is the root cause of poor outcomes, with inadequate structures and processes being the symptoms.

While reviewing an organization's strategic plan can tell you the organization's aspirations, it does not tell you what the culture is. Culture is evident in what people do, not what a plan says. When consistency is lacking in processes or performance, culture should be reviewed. Culture can sometimes be documented through a review of outcomes. This is because poor outcomes are often a result of poor culture.

For example, Tooele County School District struggled with noncompliance and low employee morale, resulting in ineffectiveness. The district performed at the bottom of its peer group across several dimensions, including average proficiency for grades 1–3, average proficiency for grades 3-8, 18+ composite ACT score, and graduation rates. This corresponded to survey results benchmarking the district's employee morale 15 percent lower than other districts previously surveyed. This is a compelling example of the relationship between an organization's culture and its achievement of core outcomes.

Though culture only exists in groups, it originates with individuals. Surveys can provide indications of culture and should be used where possible. However, surveys can be deceiving or inflated. For instance, they don't account for poor performance, as employees are unlikely to rate their own performance low. Observations, interviews, stakeholder meetings, and small group meetings may be better indicators of the actual

Though culture only exists in groups, it originates with individuals.

culture. Look for inconsistencies between what official organization documents say and what employees say.

Tip: Observe behaviors for strong evidence of what the culture actually is.

Observations are a good way to assess culture. Visit the locations where work occurs. Look for clues about culture in the language used, meeting styles, and dress codes. Observe work processes. Observed behaviors are strong evidence of what the culture actually is. To get an understanding of culture, organizations and auditors can do the following:

1. Conduct a culture survey and compare results against other organizations' survey results. Be sure to factor in any bias when analyzing the survey results.
2. Determine if leaders are held accountable for both results and behavior. Look at the oversight framework to determine if the outcomes of an organization are being aligned to performance of the leaders. Also determine if poor behavior is being corrected. This can point to leadership and governance issues. Frequently the effects of weak governance and ineffective leadership are poor culture. These need to be addressed separately.
3. Determine if the mission and strategic plan of an organization is driving decisions. Look at results, structures, and processes and determine if these are aligned to mission and strategy. This analysis exposes value alignment or a power-driven culture.
4. Interview employees to determine if they understand how their work connects to the organization's goals. Look at communication patterns and if employees are siloed. Determine if employees feel connected to the larger purpose and mission of an organization. This shows how aligned a culture is and how aligned the organization is.
5. Review how mistakes are treated. Ask questions like,
   o "Do employees feel free to innovate and make mistakes within a risk tolerance?"
   o "Do employees talk in terms of a growth mindset where opportunities are possible for everyone?"
   o "Is there a fear-based culture where employees do not feel comfortable talking about new ideas and supporting innovation?"
6. Determine if high performers who behave poorly are tolerated. Look for evidence of favoritism. Identify any inconsistencies in policy application.

When taken together, these assessment strategies highlight where culture is strong and where it can improve. When performance is subpar, culture is a good place to look for a contributing cause. A disengaged or disenfranchised workforce points to cultural issues

in the background. At the same time, a growth mindset in the culture and a focus on accountability for results can enhance outcomes.

For example, the Department of Workforce Services previously used an ineffective pay-for-performance system that promoted gaming and isolation. The department worked to fully implement 13 of 14 audit recommendations in about a year and a half. The department eliminated the old pay-for-performance system and replaced it with a financially sustainable, team-based incentive program. This rewarded collective achievement and fostered trust among employees. The department maintained existing quality and quantity standards to uphold these performance indicators. This example illustrates how positive culture can amplify an organization's efforts and results.

Proxy indicators can also help determine the general well-being of the workforce. For example, turnover rates are a good indicator of culture, particularly when reviewing against trends or benchmarks. High complaints and incidents can also signal concerns. A review of the compensation practices, including bonuses and promotions, can illustrate what organizational leaders value. Compare these to annual performance appraisals.

Because culture is multi-faceted, it is important to ensure the congruence of multiple indicators. When issues exist at the individual level, these are personnel rather than cultural concerns. However, culture can exist in pockets, so keep track of divisions and operating units when performing tests on cultural health.

Subsection 2.4
Diagnosing Structure Issues

Structure Policies Standards
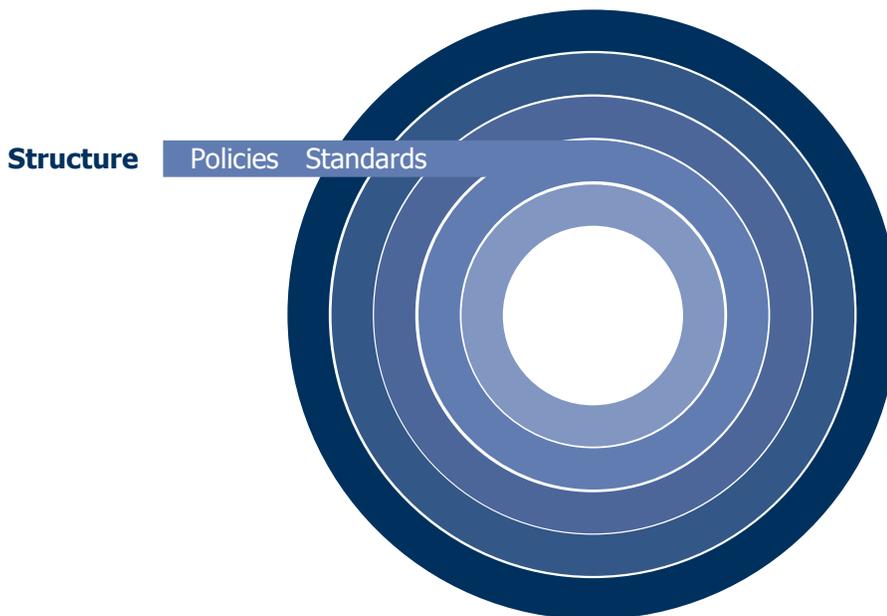
Diagnostic Checklist

Structure – The policies and standards that guide work.

Common Deficiency	Evidence to Collect	Indicators / Metrics
Unwritten policies	Interviews, Policies and procedures manual	Informal or outdated policies
Lack of standards	Standard operating procedures, Quality measures	Vague standards, % of defective products or quality concerns
Low output	Output measures	# of outputs trends
High output but unchanged outcomes	Output and outcome measures, Logic models	# of outputs relative to outcomes trends, Missing logic models
Misalignment	Utah Code, Strategic Plan, Policies and procedures manual	# of misaligned activities
Missing controls	Risk assessments, Policies and Procedures manual, audit plans	# of missing controls, misalignment between risk assessments and audit plans
Low monitoring	Monitoring logs	Frequency of review

The purpose of structure is to direct the flow of work and establish what is and is not acceptable. It is made up of the policies and standards that guide work. When structures are not formal, it can expose the organization to risks. It can also create inconsistencies in how the work is done. At the same time, overly bureaucratic environments bog down systems and undermine morale. This can impede rather than support the work. A key measure of whether

A key measure of the structure layer is the extent that policies and standards reliably produce a set of outcomes for the organization's stakeholders.

structures are appropriate is the extent that policies and standards reliably produce a set of outcomes for the organization's stakeholders.

Policies Cannot Be Assessed When Noncompliance Is Pervasive

Before determining if structures are working as intended, review compliance. If noncompliance is a widespread issue, parts of the structure are not working. This will likely be the monitoring, oversight, and policy systems, though deeper layers of cause (such as culture, leadership, and governance) could also be factors. Compliance comes first because you cannot tell if the other parts of structure are reliably producing outcomes until compliance is stable.

For example, program leadership redesigned the Ryan White program to improve resource management and program alignment with the mission. This increased compliance and fidelity. Program staff established comprehensive wrap-around medical and oral healthcare to low-income individuals living with HIV/AIDS. They also realigned funding toward these efforts. Evaluations of these new controls revealed a reduction in administrative burden, greater reach into rural areas, increased client confidentiality, and increased savings of nearly 1,000 dollars per enrolled client. By assuring implementation fidelity prior to evaluation, the program was able to successfully evaluate its impact.

Evaluating compliance is fairly straightforward. Review the rules, policies, and standard operating procedures (SOPs). Then use administrative data and observations to see if these are being followed. Another simple approach is to review instances of noncompliance for the corrective actions that took place.

Policies and Standards Should Consistently Produce Outputs and Outcomes

Once compliance has been assured, leadership should analyze what outputs and outcomes result from these structures. Analysis should aim to produce reliable information on performance. Outputs are the immediate products and services resulting from work activities. Examples include the following:

- Number of potholes reduced
- Number of smoking cessation classes taught
- Mean response time from notification to resolution
- Miles of snow removal

In theory, output measures should translate to better outcomes. If the organization engages in a lot of activity but outcomes do not improve, the structures are not working as intended.

For example, the Division of Air Quality allocated significant grant funds to the Wood Stove Conversion Program. The intended outcome was to reduce air pollution. However, our audit found that the division did not specifically target households that burn wood as the sole source of heat. As a result, the estimated reduction in emissions accounted for little of the total woodsmoke in the state.

In a positive example, decades ago the Legislature captured overhead funds used for research in higher education. These were deposited into the state's general fund as reimbursement for costs the state had already paid. Legislators noticed that the institutions were bidding on contracts with very little overhead reimbursement. This allowed the institutions to be awarded more contracts but forced the state to pay more overhead from non-grant funds. The Legislature shifted policy to allow the institutions to retain the overhead awarded, resulting in a greater incentive for the institutions to bid more. Total research grants at the University of Utah alone grew by nearly 400 million dollars over a 30-year period, in part due to this change. The audit found that the increased research activities had indirect and direct ripple effects on state and local economies.

Organizations should determine if leaders are familiar with the output and outcome measures of their programs and divisions. Determine to what extent these are used in leadership meetings and whether they inform structure recalibration. Some questions to ask include the following:

- Are outputs and outcomes tracked by each unit of government?
- Are output and outcome measures used to amend policies and standards?
- Do structures align with the organization's strategic plan and goals?
- When incidents occur, are structures adjusted to mitigate future risk?

As discussed in the leadership section, structure and leadership have a close connection. Because it is leadership's responsibility to establish and monitor the structures of the organization, miscalibrated structures reflect poorly on leadership competencies.

Subsection 2.5
Diagnosing Process Issues

Process | Series of Steps
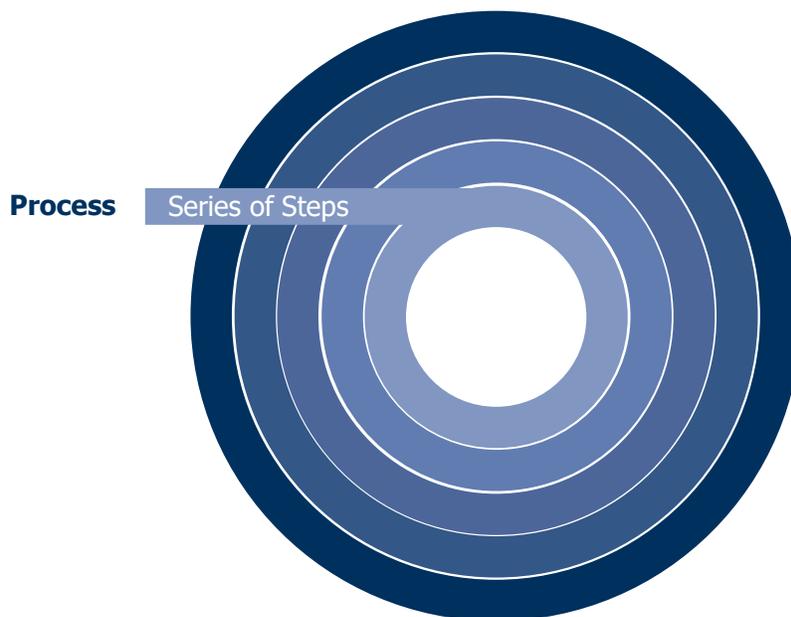
Diagnostic Checklist

Process – The steps taken to accomplish a task.

Common Deficiency	Evidence to Collect	Indicators / Metrics
☐ Lack of clarity	Process maps, Flowcharts, Interviews	% of rework
☐ Inconsistencies	Audit reports, Products/service logs	Variation in output quality
☐ Delays	Workflow logs	Average duration, % complete within expected timeframe
☐ Backlogs	Case counts, Queue time	# of cases received, Average duration in queue
☐ Low utilization	Room occupancy/machinery records	Room occupancy rate, % of time idle
☐ Resource deterioration	Downtime records	# of hours in repair, Frequency of malfunction
☐ Missing feedback loops	Performance measures	Lagging signals

The purpose of a process is to accomplish a task. Through a sequence of steps, government organizations create products and offer services. While processes are often associated with front-line work for the customer, they are universal. Processes are the logistics—the capital, workflows, timing, and practices—that move work along. In this sense, processes permeate all parts of an organization.

> **The purpose of a process is to accomplish a task. Processes tend to be visible and manageable.**

Processes tend to be visible and manageable. Because processes are task-oriented, leadership can refine processes through reorganizing and redesigning workflows. Most processes are confined to a small part of the overall work. However, when processes are linked together, they are the most concrete manifestation of work. In addition, some processes are so central to an organization's core service that fixing a single process can yield millions of dollars in savings. For example, the process for IT procurement is so sweeping that optimizing it carries the potential for millions of dollars in savings.[8]

That said, most processes are small in scope and impact. In addition, the governance, leadership, culture, and structure of the organization dictate the number of processes and their characteristics. This is why processes are the lowest level of cause.

Processes should be reviewed for common issues. These include the following:

- Processing delays
- Malfunctioning equipment
- Low quality
- Long wait times

- Costly products or services
- Frequent rework
- Inadequate resources
- Low throughput

High costs, long processing times, employee burnout, and dissatisfied customers often signal process issues. For example, the Salt Lake Valley Emergency Communications Center consistently failed to meet emergency call standards of timeliness. Over 35,000 calls were not answered in a timely manner in 2019 alone. This led to widespread wait times for emergency callers. To address this, the audit recommended better coordination and increased monitoring and training.

Common reasons for process issues include the following:

- Undocumented steps
- Unclear ownership
- Missing quality standards
- Improper sequencing
- Inadequate tools

- Task overload
- Under-utilization
- Broken handoffs
- Missing signals
- Inadequate training

Measure the time it takes to complete each step of a process. Look for wait times where a product or service has stalled. Be alert to "hockey stick" graph patterns, which suggest work is postponed until it nears timeliness requirements set in rule. Determine the frequency of rework.

---

[8] See "Low-Cost IT Procurement Efficiency Evaluation", 2022, Office of the Legislative Fiscal Analyst and Governor's Office of Planning and Budget.

Staff interviews are insightful as to whether the right tools are available and if workloads are manageable. Leadership can verify these interview insights by observing employees as they use tools and by analyzing caseloads.

For example, the Division of Adult Probation and Parole experienced low compensation and heavy workloads, leading to burnout and diminished service quality. Police departments competed with the division on pay, leading to high turnover rates. The audit recommended increasing supervisor reviews of agent caseloads. We further recommended that the division create and implement retention strategies to ensure adequate staffing of experienced agents.

Leaders should review data on output quantity and quality. Determine if product defects are frequent or substantial. Perform a cost-benefit analysis to relate costs to the number of outcomes. To understand customer experience, refer to the Governor's Office of Planning and Budget customer experience reports. These can help leaders and auditors understand customer satisfaction or complaints.

See the **Customer Experience Reports**.

The final section of this handbook shifts from the causes of issues to the effects of interventions. These methods will help organizations determine the impact of their efforts.

The Best Practice Handbook:
Diagnosing Root Causes and Driving Results (Open comment Period)

Section 3
Cause Is the Key to the Solution

Cause tells leaders where and how to intervene. A doctor relies on their diagnosis of disease to prescribe an effective treatment. Similarly, leaders must use the knowledge gained from root cause analysis to intervene effectively on organizational issues.

Interventions aim to change the conditions causing negative outcomes. Root cause analysis identifies which conditions need to change. By accurately identifying and treating the causes of organizational issues, leaders proactively eliminate them. This contrasts with a reactive approach, where leaders are continuously putting out fires.

To ensure issues improve, leaders should monitor outputs and outcomes.[9] Causal reasoning tools further confirm interventions are effective. Leaders should apply these methods to sustainably resolve organizational issues.

Interventions That Target the Causes of Issues Are More Successful

Interventions that target causes often achieve the best return on investment. In addition, solutions tend to last longer because they move beyond treating symptoms. The main purpose for identifying the causes of issues is to increase the effectiveness of interventions.

As discussed in Section 1, leaders should intervene at each of the five layers where they identify a cause. In the example of the Utah Schools for the Deaf and the Blind (USDB), the audit team recommended that leaders improve statewide plans (structure) and enhance performance tracking (leadership). They further recommended that the Legislature consider an alternate governance structure (governance). These recommendations addressed the identified causes of the performance issues.

The following is a case study that demonstrates how interventions should address the underlying causes of organizational issues. It also highlights the critical relationship between leadership and culture.

---

[9] *The Best Practice Handbook: A Practical Guide to Excellence for Utah Government* (Report No. 2023–05), p. 72. Office of the Legislative Auditor General.

The Utah Department of Corrections Demonstrates the Power of Treating the Causes of Issues

The Utah Department of Corrections (UDC) experienced serious safety issues following the move to the new state correctional facility in 2022. Our audit report[10] noted that correctional officers (CO) and inmates were at greater risk of death or injury within the prison due to staffing constraints. One assault incident highlights the severity of the situation:

> **Maximum-Security Assault:**
>
> *In the maximum-security building (Antelope), a single CO entered the highest-level security section, alone. He requested that all sub-day room doors in the section be open at the same time as he did a routine cell check. As he stepped into a sub-day room, an inmate who had not been properly returned to his cell was hiding in the shower of his sub-day room and assaulted the CO from behind. The inmate handcuffed the CO. No other officers were present to help, and there was no response for two minutes, well after the assault had progressed to life-threatening levels despite being visible on security cameras.*

The issues of safety stemmed from a persistent challenge to maintain adequate staffing levels. Our audit estimated that full staffing needs were 895, but UDC only employed 323 actual staff. This created a shortage of 572 staff.

The staffing issue was a symptom of several deficiencies across the five layers of cause framework, including the following:

- **Governance**—Increased demand for certified staff due to prison design

- **Governance**—Misaligned financial incentives for officers, undermining retention

- **Leadership**—A lack of focus on safety and security needs in the strategic plan and inadequate emphasis on hiring and retention

- **Leadership**—According to staff survey responses, a lack of senior leadership striving for excellence and innovation

- **Leadership**—According to staff survey responses, a lack of leadership engagement, skill, and innovation, negatively affecting culture[11]
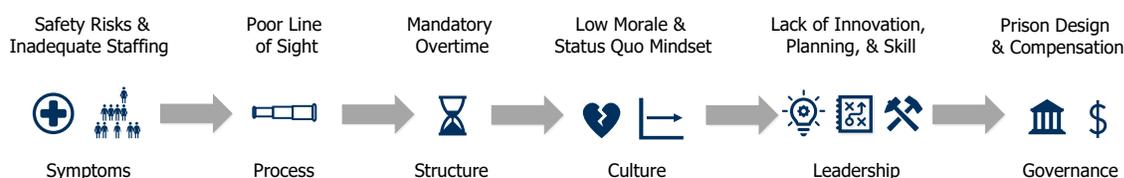
---

[10] *A Performance Audit of the Utah State Correctional Facility: An Examination of Staffing, Culture, Safety, and Security* (Report No. 2023–17), Office of the Legislative Auditor General.
[11] Almost 1 in 5 staff favored a change in leadership.

- **Culture**—Low morale, leading to absenteeism

- **Culture**—A status quo mindset, leading to unresolved staffing issues

- **Structure**—Mandatory overtime policy, leading potentially to burnout and officer turnover

- **Process –** Poor line of sight, increasing staffing demands

As evident in this list, several causes negatively influenced staffing and safety at the prison. Figure 3.1 illustrates the set of causes contributing to negative outcomes.

Figure 3.1. The Utah Department of Corrections Struggled with Safety and Staffing Issues. Several causes contributed to the issues, making them difficult to overcome. In particular, leadership issues negatively affected culture, undermining the intended outcomes.

Safety Risks & Inadequate Staffing	Poor Line of Sight	Mandatory Overtime	Low Morale & Status Quo Mindset	Lack of Innovation, Planning, & Skill	Prison Design & Compensation
Symptoms	Process	Structure	Culture	Leadership	Governance

Source: Auditor generated.

Our audit made recommendations across all five layers of cause to address these serious concerns. The audit found that leadership failures significantly contributed to the culture of low morale and a status quo mindset. These leadership and cultural issues rippled through the organization, leading to downstream effects, such as stuck processes and ineffective policies. It also had upstream effects as leadership did not always adequately work with the Legislature. These issues led to funding and governance concerns and staff feeling unsupported. A lack of safety inside the prison led to stress and low morale. In short, effective leadership and positive culture appeared to be the missing linchpins in the production of better outcomes.

Strong Leadership Can Improve Culture

Just before the audit's release, a new executive director took over the department. The new leader concentrated on improving the culture of the organization by making critical changes across the five layers of cause, including the following:

- **Governance**—Worked with the Legislature to realign compensation
- **Leadership**—Analyzed staffing capacity to support planning
- **Leadership**—Planned space utilization, optimizing the use of resources
- **Leadership**—Established new career paths, incentivizing engagement
- **Culture**—Focused on safety as the top priority, improving morale
- **Culture**—Included staff in strategic planning, promoting buy-in
- **Culture**—Met with line staff to understand their issues

- **Culture**—Promoted innovation to encourage a new mindset
- **Structure**—Broke down bureaucratic barriers, leading to change
- **Process**—Observed and streamlined steps for completing tasks

While the new leader intervened at each of the five layers of cause, every change had the effect of supporting staff. By focusing on culture, the leadership team amplified its influence. As staff felt more supported and included, engagement and collaboration improved. The audit team had observed during the audit an "us" versus "them" mentality between staff and management. The new leader changed this to a "we" mentality that advanced the mission and core outcomes of the department.

Data we received from UDC in November 2025 showed that UDC made gains in a short amount of time. Within 2 years:

- Inmate assaults decreased
- No officer assaults occurred in three months of a 12-month period
- Contraband detection was enhanced
- Staffing levels greatly improved

This example highlights the importance of strong leadership in bolstering a positive and collaborative culture. As culture improved, issues resolved and outcomes improved.[12] Causes at all five layers were undermining results and needed correction. New leadership demonstrated a commitment to building a good culture as the means to better outcomes. Meaningful improvement followed.

As discussed in Section 2, leadership sets the tone at the top of the organization. This affects the organization's culture. In the case of an ineffective leader, this can undermine employee engagement and, ultimately, outcomes. On the other hand, an effective leader can positively shift the culture of the organization and use their position to petition necessary changes. Consequently, we believe that the selection and development of effective leaders can have an outsized effect on organizational success. In turn, this translates into better outcomes for the citizens of the state.

Not all issues have been resolved at the prison. However, we believe the presence of an exceptional leader who focused on promoting a positive culture created the conditions necessary for the desired outcomes to occur. The next section describes causal reasoning tools that further help leaders determine whether the intervention was the cause of improvement and by how much. This enables leaders to focus limited resources on what works.

---

[12] UDC executive management attribute success to ongoing investment in their strategic pillars: safety, staff investment, innovation, collaboration, reentry, and rehabilitation. We believe strong leadership and improved culture were central to the success of these initiatives.

Causal Reasoning Tools Amplify Impact

We want the best solutions to issues. Causal reasoning tools determine what the best solutions are. They do this by isolating the effects of an intervention.[13] This allows us to determine its true impact. We encourage leaders to use these tools wherever possible to improve the effectiveness and efficiency of their interventions. Because these methods are more technical in nature, the agency may need to partner with a university or other entity that has the necessary expertise.[14] In addition, the Office of the Legislative Fiscal Analyst (LFA) and the Governor's Office of Planning and Budget (GOPB) have staff that can review pre-analysis plans and offer feedback and support.

To maximize value, leaders should carefully plan and evaluate an intervention. Well-designed interventions can provide information on what does and does not work. Leaders need to know how much the intervention caused outcomes to improve. This enables them to pivot away from unsuccessful solutions and expand productive ones. The results are

- Better outcomes
- Lower costs
- Enhanced decisions
- Healthier organizations
- Advanced missions
- Increased public trust
- Greater impact

When leaders are not carefully planning and evaluating interventions, organizations are vulnerable to persistent issues, ineffectiveness, and waste. While there are many causal reasoning tools that leaders use to assess the impact of an intervention, three common tools are

- Randomized Control Trials (RCT)
- Difference-in-Differences
- Regression Discontinuity

All three causal reasoning tools we describe rely on the foundation of comparison. The central idea is that by comparing outcomes between two groups—one that undergoes

---

[13] An intervention is an attempt to change the conditions causing a negative outcome. The goal is that this change causes outcomes to improve.

[14] In Appendix A, we provide templates for data sharing and data use agreements. Agencies can modify these templates to suit their needs when establishing partnerships. Any costs associated with the analysis should be established upfront, along with a clear pre-analysis plan. See Appendix B.

an intervention and one that does not—we can gain knowledge of the true impact the intervention had. Because both groups exist at the same time, other influences (such as the economy or policy changes) can largely be ruled out as the causes of outcomes. We encourage leaders to use these tools wherever possible to improve the effectiveness and efficiency of their interventions.

Randomized Control Trials Are the Gold Standard For Determining the Impact of an Intervention

The gold standard for creating comparison groups is the randomized control trial (RCT).[15] In an RCT, the intervention is rolled out to a treatment and control group simultaneously. The group that receives the intervention is the treatment. The group that does not receive the intervention is the control. Because both groups exist at the

To determine cause, we need to know what would have happened were it not for the intervention.

same time, we can conclude that the intervention caused any differences in outcomes.

Treatment (Intervention)

Control (No Intervention)

Randomization—a key feature of an RCT—ensures that each of the groups is sufficiently similar.[16] This eliminates bias in the selection process. It ensures that both known and unknown characteristics of the groups are, on average, not unfairly influencing the outcomes.

---

[15] There are many ongoing government efforts to use RCTs to assess program and policy impact. For example, the Foundations for Evidence-Based Policymaking Act of 2018 is a federal statute that supports more robust evaluation. The Evaluation.gov website contains many useful resources related to RCTs, including an Evaluation Toolkit. Other helpful resources include The Office of Evaluation Sciences, Federal Data Strategy, Federal Enterprise Data Resources, J-PAL and Better Evaluation. These resources can be used by policymakers and budgeting professionals to rigorously evaluate the effectiveness of programs. Leaders can also use these resources to guide evaluation of interventions.

[16] A sample size of 200 or more is usually large enough to balance group differences.

Figure 3.2 illustrates the RCT design.

Figure 3.2. Randomized Control Trials Amplify Impact by Determining Whether the Intervention Works. An RCT splits participants into two groups. One group does not receive the intervention while the other group does. Because both groups exist at the same time, we can conclude that the intervention caused any differences in outcomes.
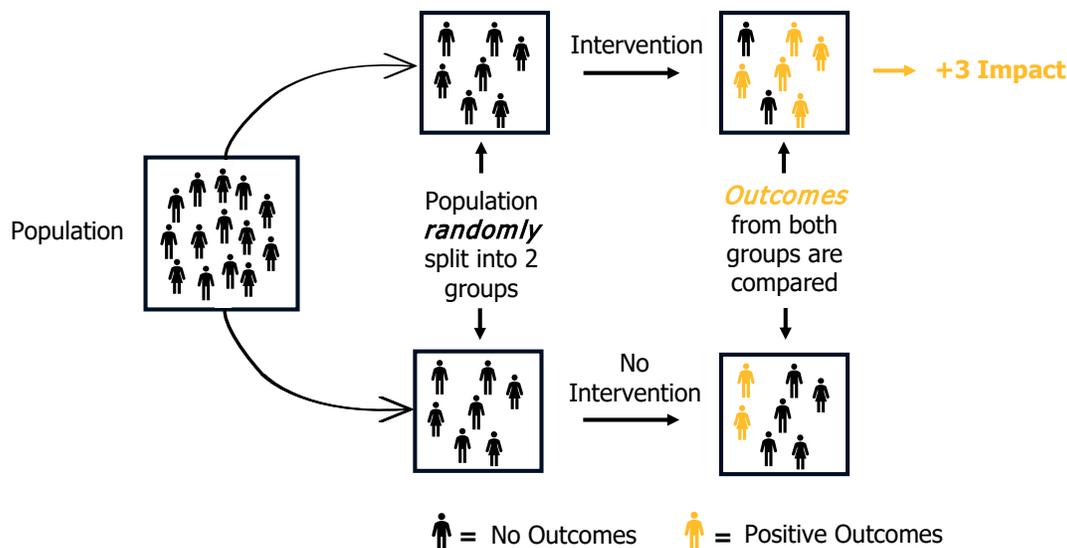
RCTs are a powerful tool for determining how well an intervention works. The value is leaders can use the knowledge gained from the RCT to decide whether to expand the current strategy or try a new one. This focuses limited resources on strategies that are effective, reducing waste.

> **Leaders can use resources found at the Evaluation.gov website to help them evaluate the impact of interventions.**

Returning to the prison example, we can explore how RCTs could have been used to ensure interventions were having the intended effect. This ensures leaders maximize their impact. Some potential RCTs include the following:

- **Culture RCT:** Leaders could randomly assign work groups to participate in an employee wellness initiative. Outputs and outcomes to compare include absenteeism and morale, as measured in survey response scores.

- **Leadership RCT (Retention):** Leaders could randomly assign staff to participate in a mentorship program. Outputs and outcomes to compare include officer turnover rates, absenteeism, and morale, as measured in survey response scores.

- **Leadership RCT (Hiring):** Leaders could randomly assign job applicants who meet minimum qualifications to alternative hiring processes, such as streamlined

or multi-stage interviews. Outputs and outcomes to compare include three-year retention rates.

Once an RCT is complete, leaders compare outcomes for both groups.[17] For example, after implementing the mentorship program, leaders compare absenteeism for the control and treatment groups. If absenteeism for the control group averages 5 more days a month than the treatment group, then leadership can conclude that the mentorship program caused a reduction in absenteeism by that amount. Other outputs and outcomes could also be compared, such as morale and turnover rates. Leaders should use these findings to make decisions about the value of the intervention. This includes whether to expand or pivot to a new strategy.

Randomized Control Trials Should Only Be Used in Certain Situations. While RCTs do provide powerful information on the impact of government interventions, they are not always possible. For example, certain benefits programs are designed around eligibility thresholds, and it would not be ethical to randomize groups to determine who receives the benefit.[18] Because RCTs require a design element, policies and programs that already exist cannot be evaluated using an RCT unless undergoing a change. In these situations, other techniques are more useful. Several statistical techniques are designed to deal with these situations and can still provide a good approximation of impact. Two of these are discussed in the next subsection of this report.

Despite the limitations that exist, RCTs are a practical method that can be used in numerous scenarios. Many interventions targeting the causes of organizational issues are well suited to an RCT approach. Similarly, new pilot programs are a natural place to incorporate RCTs.

Specialized Methods Can Be Used When an RCT Is Not Feasible

When an RCT is not feasible, leaders should determine if an alternative approach can be used to assess impact. The difference-in-differences and regression discontinuity methods are common tools for imitating the logic of the RCT. We are currently working on developing a project portal to facilitate such connections. Our office also offers

---

[17] While a full discussion of RCTs lies outside the scope of this report, it is important to note potential biases that can enter during implementation. For example, "spillover effects" refers to the treatment group influencing the control group by sharing information or practices about the intervention. Other potential biases include non-participation, dropout rates, blinding, and logistical challenges. Leaders should be mindful of these pressures when designing and implementing an RCT.

[18] Ethical considerations are critical. However, Dr. David Yokum explains that because the effectiveness of an intervention or new program is typically unknown, many of the ethical considerations surrounding treatment are not an issue.

training on causal analysis and causal reasoning to build capacity within the government workforce to effectively carry out these types of studies.

The Difference-in-Differences Method Works when a Good Comparison Group Exists. A good comparison group means the output or outcome has followed a similar trend in another location, as seen in Figure 3.3.

Figure 3.3. The Difference-in-Differences Method Identifies a Good Comparison Group by Looking at Output or Outcome Trends Over Time. The parallel trend suggests that the treatment group would change in similar ways were it not for the treatment. This allows us to determine the effects caused by the intervention.
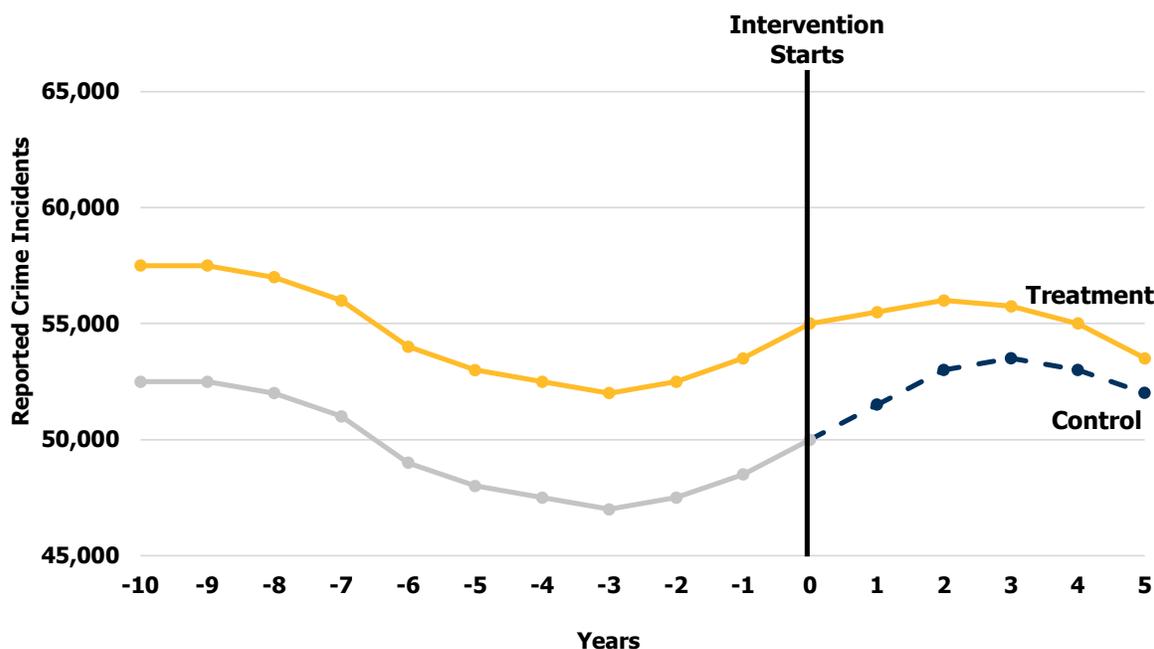
Source: Modified from graph provided by Dr. David Yokum, Chief Scientist, North Carolina Office of State Budget and Management.

The parallel trend suggests that the treatment group would continue to change in similar ways to the control group were it not for the treatment. As

done in the RCT, the output or outcome of the control group is compared to the treatment group after implementation. This allows us to estimate the effects caused by the intervention.
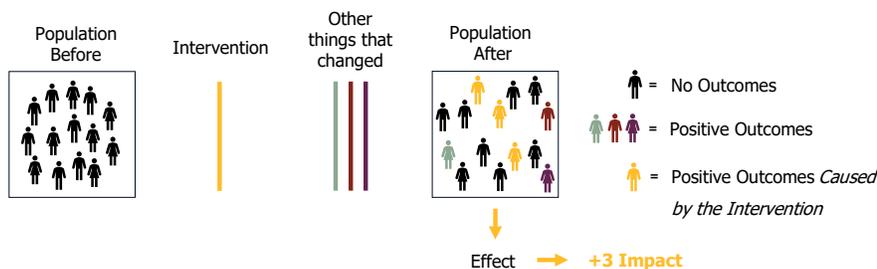
Returning to the prison example, leadership could use a difference-in-differences study to evaluate an intervention that is more difficult to implement through an RCT.

- **Governance Difference-in-Differences**—Leaders could group prison units or other state prisons serving similar population types. Within each group, leaders could compare units that implement the fully staffed direct supervision model to units that continue operating under reduced staffing levels. Groups must demonstrate parallel trends prior to the intervention. Outcomes to compare include assault rates for both groups.

While not as straightforward as an RCT, the difference-in-differences approach can inform leaders of likely success before implementing an expensive intervention. This is because it provides evidence of an intervention's expected impact. It does this by comparing similar groups over time to control for other influences.

The Regression Discontinuity Method Works When a Program Has an Eligibility Threshold. This is common in benefits programs where criteria are used to determine who receives a benefit. It also is common for academic or training programs using qualifications for sorting into groups. In these situations, groups just above and below the threshold are used to compare outcomes. Because the eligibility threshold is already a natural part of the program, there are reduced ethical considerations.

Figure 3.4 shows a graph of the regression discontinuity approach.

Figure 3.4. Groups Just Above and Below the Eligibility Threshold Are Used to Compare Outcomes. Because the eligibility threshold is a natural part of the program, there are reduced ethical considerations. Outcomes for both groups are compared to determine the average impact.
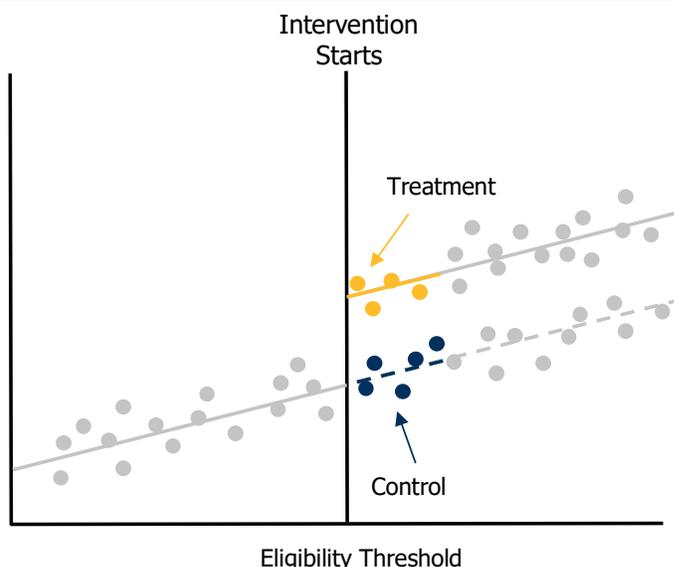

Source: Modified from graph provided by Dr. David Yokum, Chief Scientist, North Carolina Office of State Budget and Management.
* The dotted line indicates that the data is unobserved.

By limiting cases in the two groups to those close to the eligibility threshold, we establish their similarity. The group that just barely missed the cutoff for eligibility into the program is used as the control. The group above the cutoff is the treatment group. The differences in outcomes can again be attributed to the impact of the program.

To again return to the prison example, senior leadership could use a regression discontinuity to assess a leadership training program within the prison.
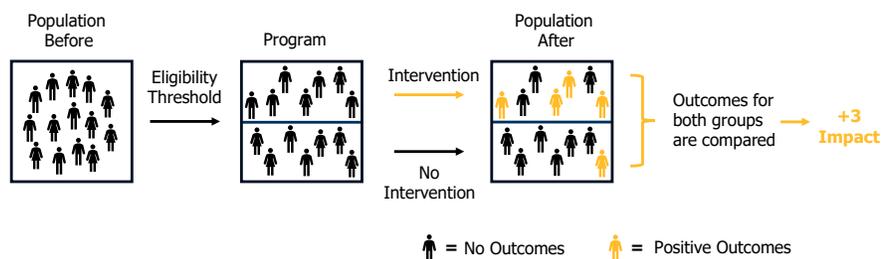
- **Leadership Regression Discontinuity**—If leaders introduced a new, intensive rehabilitation program at the prison, participation could be based on risk assessment scores. Those below the cutoff score would act as the control group, and those above the cutoff would be the treatment group. Outputs to compare include disciplinary actions between groups.

As mentioned earlier, leadership may need to do some technical work to balance the two groups during the analysis phase. For this reason, it may be necessary to work with outside experts to ensure the analysis is set up and executed correctly. Seemingly small errors can have an outsized effect in the findings and conclusions of the analysis.

Without implementing a causal reasoning tool, it is easy to draw the wrong conclusion. Causal reasoning tools avoid giving credit to interventions that seem effective when they are not. They also avoid assigning blame to interventions that seemed ineffective, when they were actually impactful. Importantly, causal reasoning tools also tell us the size of the impact. This enables leaders to refine strategies to optimize efficiency. It also empowers leaders and policymakers to make essential cost-benefit decisions.

When planning to use any of the causal reasoning tools, it is best practice to create a pre-analysis plan.[19] A pre-analysis plan accomplishes several things:

- Ensures the methodology selected will produce reliable information
- Builds consensus on what the program is seeking to influence
- Establishes agreement about how to interpret results of the evaluation
- Sets the scope of the evaluation, including timelines and costs

Because an intervention is attempting to cause new results, leaders should establish the expected timeline for outcomes to be achieved. Some outcomes, such as absenteeism rates, could be evident quickly. Other outcomes, such as retention rates, could take years to occur. Clearly defining when the results are expected to be achieved enables evaluators to accurately determine impact.

Better Evaluation Means Better Results

As discussed at the beginning of this report, better evaluation is a key leverage point for improving planning and implementation. As evaluations improve their relevance and reliability, decision-makers can take meaningful action to improve organizations.

Root cause analysis and causal reasoning are two sides of the same coin. The first assists organizations in identifying the causes of issues. The second guides effective resolution of those issues. Both promote effective and efficient government, which is always the goal. Our office will release additional resources and tools to further support organizations as they work toward building capacity for root cause analysis and causal reasoning. We encourage leaders to reach out to our office, LFA, and GOPB for

Request training on root cause analysis or causal reasoning from OLAG.

---

[19] A detailed guide for creating a pre-analysis plan can be found in Appendix B of this report.

support as they strive to promote better outcomes for the citizens of the state of Utah through better evaluation.

Appendices

**A. Data Use Agreement Template
Source: Dr. David Yokum, Chief Scientist,
North Carolina Office of State Budget and Management**

The Best Practice Handbook:
Diagnosing Root Causes and Driving Results (Open comment Period)

**DATA SHARING AGREEMENT
BETWEEN [DATA PROVIDER] AND THE [DATA RECIPIENT]**

This Data Sharing Agreement ("Agreement") is entered into between [Data Provider] ("[Acronym or Abbreviation]") and [Data Recipient] ("[Acronym or Abbreviation]") (collectively referred to herein as the "Parties" and each individually referred to herein as a "Party"), as of the date of the final authorized signature on this Agreement.

1.   PURPOSE OF THE AGREEMENT

  a.   The Parties seek to better leverage empirical insights and methodologies in order to improve public policies and programs. [Data Recipient] will provide expert services related to the curation and usage of administrative data, owned by [Data Provider], for research purposes.

  b.   To facilitate the use of administrative data for research purposes, [Data Recipient] will create and curate an Integrated Data System ("IDS"). The IDS will entail agency datasets linked at the individual level across agencies, geographies, and time, stored within a secure computing environment and including protocols for using the IDS (e.g., security procedures and statistical software). [[Data Recipient] will also create commonly available resources (e.g., data dictionaries) and cultivate relationships between academia and government, to support the generation of actionable intelligence from the IDS.]

  c.   The purpose of this Agreement is to set the terms by which [Data Provider] will physically transfer its administrative data to [Data Recipient], as part of participating in the IDS. It also authorizes a narrow set of activities related to cleaning, documenting, and linking the data with other data within the IDS.

  d.   This Agreement does *not* authorize the usage of any part of the data referenced herein for research. No reports or publications on policies and programs associated with the data can be created. Separate data use agreements are required to authorize such usages.

2.   SCOPE OF WORK

  a.   The scope of work is restricted to only those activities a) required to clean, document, and link the data—as defined herein—as part of the IDS; b) to conduct work that improves the data's quality for research purposes; or c) explains the methodologies used to accomplish these activities.

       i.   Data cleaning is the process of detecting and correcting (or removing) corrupt, inconsistent, or inaccurate records from a data set.

1.   <u>Example</u>: Data stored in different formats (e.g. January 1, 2019 versus 1/1/19 versus 01.01.2019) may be converted into a single standardized format.

ii.   <u>Data documentation</u> refers to written descriptions of the data needed to ensure the data can be understood and interpreted by users.

1.   <u>Example</u>: This may include information identifying the original source of the dataset and the date range covered; or a data dictionary could list all the variables in a dataset and describe what those variables mean.

iii.   <u>Data linking</u> refers to techniques of connecting pieces of information that are thought to relate to the same person, family, place or event.

1.   <u>Example</u>: If Jane Doe has a record in an education dataset and there is another Jane Doe record that exists in a health dataset, a linkage could connect the two records under the assumption it is the same Jane Doe in both datasets. Some data linkages must be inferred. For example, if there is a Jane Doe and a Jane B. Doe, but both records share the same address and birthday, we could reasonably assume it is the same person, it just happens that the middle initial was not included or dropped from one of the datasets.

iv.   The "<u>to otherwise conduct work that improves the data's quality</u>" clause is to acknowledge that activities beyond more straightforward forms of cleaning, documentation, and linking may be valuable in enhancing the quality of the data. Some such activities may even be a type of research

1.   <u>Example</u>: It will likely be valuable to compare a variety of automated linking techniques with hand connected linkages, to learn which automated technique is most accurate. This is an activity that could, in itself, be published as a methodology paper. The intention here is to acknowledge, permit, and even encourage such work—since it will enhance the quality of the IDS—but to make clear that, in order to pursue a public facing report or paper about that work, a DUA would be required.

b.   The prohibition against using data for reports or publications absent a DUA is not meant to prevent the use of internal memos amongst users authorized to engage in the activities described in this section. Staff working to clean, document, link, or otherwise conduct work that improves the data's quality

will likely need to write memos to each other describing ideas, methods, and results of undertaken work. Consideration of the memo's purpose provides a guide: a memo to facilitate work to enhance the quality of the data would be permissible under the DSA, while a memo to facilitate work about the policies or programs related to the data would constitute a use requiring a DUA.

3.   DATA SHARING

a.   Within five (5) business days after the last Party signs this Agreement, [Data Provider] will transfer to [Data Recipient] the data listed in Appendix A (the "Data") into the [NAME OF INTERGRATED DATA SYSTEM]

b.   Appendix B describes methods for transferring the Data into Stronghold, including usage of the secure transfer server.

c.   The Parties will work together to share the Data in a format most amenable for integration into the IDS and for use in research.

d.   Agency will share any relevant documentation about the Data (e.g. data dictionaries) that could be useful to its integration into the IDS and for use in research.

e.   If the Data is derived from a database, then Agency will share the exact query code used to pull the Data from the database.

f.   The Parties will coordinate to determine a reasonable frequency by which to transfer updated and new data. It is recognized that some Agency data may be updated and created more frequently than other Agency data, and thus the reasonable frequency of sharing updated data is likely to vary depending on the dataset in question. Particular uses (covered by an appropriate DUA) may also trigger the need for a more frequent data transfer. As a baseline, however, the Parties intend to share data monthly.

4.   APPLICABLE LAWS

a.   [This section includes the statutory authorities that grant the Parties the ability to share the data shared under this Agreement and specify the data security requirements each Party must fulfill under this Agreement.

b.   Refer to the Applicable Laws Section Instruction Sheet to fill in this section]

c.   Sensitive data is information to be protected against unwarranted disclosure. Protection of sensitive data may be required for legal or ethical reasons, for issues pertaining to personal privacy, or for

proprietary considerations. Sensitive data includes, for example, personally identifiable information.

d. <u>Personally identifiable information</u> ("PII") is information that can be used to distinguish or trace an individual's identity, either alone (e.g. full name or social security number) or when combined with other personal or identifying information (e.g. date or place of birth) that is linked or linkable to a specific individual.

5. DURATION OF AGREEMENT

a. The "**effective date**" is the date when the last party signs this agreement. This agreement will continue for five (5) years from the effective date.

b. If mutually agreed by the Parties, the agreement can be renewed in five (5) year increments.

6. TERMINATION OF AGREEMENT

a. ADD

7. DATA CUSTODIAN

a. Brown designates [name], [title of individual], as the Data Custodian.

b. The Data Custodian is the Brown employee or agent who is responsible for receiving the Data from Agency.

c. Once the Data is received from Agency by the Data Custodian, the Data Custodian is responsible for:

i. Providing the Data to authorized users and ensuring that authorized users receive access to the Data only in conformity with this Agreement;

ii. Maintaining a record of all Data requested and received;

iii. Maintaining a record of all submitted data use agreement requests and their status, including whether rejected or approved, and if approved the start and end dates of the DUA.

8. REPORTING UNAUTHORIZED DISCLOSURES OR MISUSE OF DATA

a. If Brown discovers any actual or suspected use, access, or disclosure of Data not authorized by this Agreement, Brown willll promptly notify Agency in writing. Such notification will include the following information (to the extent known by Brown):

i.   The nature of the unauthorized use, access, or disclosure;

ii.   The Data used, accessed, or disclosed;

iii. The individual who made the unauthorized use or access or received the unauthorized disclosure;

iv.   Brown's actions to mitigate any negative impact of the unauthorized use, access, or disclosure; and

v.   The corrective action Brown has taken or will take to prevent future similar unauthorized use, access, or disclosure.

b.   Brown will be responsible for making reasonable efforts to eliminate or mitigate the negative impact of any unauthorized access, use, or disclosure of Data. Brown will inform all unauthorized individuals who received or accessed Data that such Data was disclosed or accessed in error and of the steps the individual must take to mitigate any negative impacts of the error.

9. PARTIES' REPRESENTATIVES & NOTICES

a.   The following individuals are the Parties' representatives under this Agreement:

[Data Provider]
Name: [Contact person name]
Title: [Title]
Address: [Address]

Email: [Email]
Phone: [Phone]

[Data Recipient(s)]
Name: [Contact person name]
Title: [Title]
Address: [Address]
Email: [Email]
Phone: [Phone]

b.   All writings and notices delivered under this Agreement shall be provided to the Parties' representatives by electronic mail.

c. The Parties' representatives may be changed by a Party at any time by giving the other Party notice by email of the new representative's name, title, and contact information.

10. MODIFICATIONS

Except as otherwise stated in this Agreement, the terms and conditions of this Agreement may be modified only upon the agreement of the Parties. A modification must be in writing and signed by the duly authorized signatories of the Parties.

11. COUNTERPARTS

This Agreement may be executed in counterparts, each of which shall be deemed an original and all of which shall be taken together and deemed to be one instrument.

12. NO THIRD-PARTY BENEFICIARIES

No individual or entity shall have any right, interest, or claim under this Agreement, or be entitled to any benefit under or on account of this Agreement, as a third-party beneficiary or otherwise.

The Parties, by the signatures of the authorized representatives below, hereby acknowledge and agree to the terms and conditions of this Agreement.

IN WITNESS WHEREOF, the Parties have executed this Agreement as of the dates set forth below.

[DATA PROVIDER]

_____   _____

By: [Name]   Date
Title: [Title]

[DATA RECIPIENT]

_____   _____

By: [Name]   Date
Title: [Title]

B. Pre-Analysis Plan Template
Source: Dr. David Yokum, Chief Scientist,
North Carolina Office of State Budget and Management

Pre-Analysis Plan

Does X affect Y in direction Z?

Authored by <mark>INSERT</mark>[1]

Date of Document

Table of Contents

A. Purpose of a Pre-Analysis Plan

This plan aims to promote scientific research integrity by reducing researcher discretion after realization of experimental outcomes as well as boster political integrity by negotiating value judgments in advance (see generally Yokum & Bowers, 2022). The model of this plan comes from best practices for conducting and pre-registering field experiments.[2] We pre-commit to reporting all the analyses specified below. If we report any analyses in addition to those described here, we will indicate that they were not pre-registered.

B. Program & Project Context

A paragraph or so of a program we're trying to improve. Not necessarily the *component* of the program we're trying to improve but the bigger picture. Followed by a bulleted list of:

- Key Performance Indicators of the program;
- Insure as many Kentuckians as possible;
- Provide resume help;
- Deliver timely food benefits to the people of Indiana.

A paragraph or two calling out which KPI(s) this experiment is going to work to improve and gesturing at why we, in theory, think this experiment might improve it.

A short description of the program's timeline (e.g., "started in 2000, suspended during the pandemic, recently reinstated). And ending with a short description of *our experimental component's* timeline.

The call to action: Why do we need to do an experiment? Can be as simple as "No evidence has been gathered on the effectiveness of this program at improving the above KPI" or something similar.

B.1 Objectives

One sentence summary of the study universe, e.g., "those who claim unemployment benefits," or "people making between X and Y and thus are eligible for subsidies on the healthcare marketplace." Followed by a bulleted list describing precisely how the KPIs above will be measured, e.g.,

- Percent difference in annual earnings,
- Percentage change in insurance uptake,
- Absolute reduction in the number of drop outs.

For explicit definitions of all these terms, see the subsection titled Metrics under our Methods section.

---

[2] CONSORT Transparent Reporting of Trials. Available Consort - CONSORT 2010. Accessed 30 June 2016. See Consort - CONSORT 2010; and Gerber, Alan S. and Donald P. Green. (2012). Field Experiments: Design, Analysis, and Interpretation. New York, NY: W.W. Norton & Company.

B.2 Research Context

This section contains your primary narrative. Should tell both the history of the program you're studying (at the federal level as well as at the state/agency level) and in particular should focus on the *most recent changes* to the program as that is what we are ultimately measuring. Generally would expect some citations to the US Code or the Code of Federal Regulations or equivalent. This is also a chance to incorporate any other studies of this or similar programs.

B.3 Research Insights

This section describes exactly which decisions the results of this research will be used to inform, preferably with *deadlines*. Could be something like:

- The outcome of this study will help determine whether XYZ program is recommended for renewal during FY24;
- The outcome will guide who is eligible for this program starting in PY25;
- The outcome will shape whether to focus on in-person versus online services starting at the conclusion of this study, though preferably before contract renewal in December 2024;

C. Methods

The following section details the randomization process and subsequent analysis used in this evaluation.

C.1 Overview

Describe with specific numbers and symbols the process by which people will be enrolled (or not) in the experiment. Consider in particular any clusters implicit in the design and the frequency of selection. Place this into context utilizing historical numbers from the program if available. Flow charts may be useful here.

By the end of this overview, the reader should have a general idea of:

- How many people are likely to be in the control and treatment
- How many treatment arms there are
- The general timeline of the experiment
- When people will be enrolled in the experiment
- Any foibles of the design, e.g., "everyone is eventually enrolled"

C.2 Randomization and Selection

This section should describe in detail the process by which people will be enrolled into treatment. Specificity is greatly appreciated. Some things that should be present:

- *Who* is doing the randomization?
- *What* is the randomization algorithm (note that it should be *deterministic!*)?
- *When* do people get randomized?
- *Where* does the agency learn the information necessary to randomize the participants?
- *How* is selection communicated (or not) to the participants?

Metrics

This section should describe in exceptional detail *exactly* how the measures you described in the Objectives section will be computed. Terms like "Annual Earnings" or "Health Insurance Enrollment" or "Gainful employment" should be explicitly defined and the data from which they are computed should be enumerated. Any known foibles with the data or the computation should be explained here.

By the end of this section, any two people who were asked to compute the measures described in this section from the data described in this section should be able to do so and come to the same conclusion. This is especially important for field experiments since it will likely be awhile before you return to this project.

Helpful hints:
- **Bold** words that you are defining.
- Give explicit examples of computations, especially tricky ones that involve aggregating aggregate data (e.g., "quarterly earnings") relative to discrete cutoff (e.g., "the exact day someone lost their job")

Timeline

This section should answer the following questions:
- How long will the experiment last?
- How long will participants be followed?

Participation

This section should address potential non-compliance issues and how and whether they will be mitigated. Some examples to consider:

- For an encouragement design involving postal or electronic mail, you're likely to have many incorrect addresses
- For program enrollment, many people are likely not to show up for the program, forget to renew, or otherwise not complete the program
- For rolling designs, e.g., for training experiments, people are very likely to move their treatment date. Usually this will get moved later, but not always
- In many designs, people can refuse to be enrolled in the treatment. How frequently does this occur in practice?

Two categories of measurement

In most experiments, you will end up primarily measuring an *intent-to-treat* (ITT) effect. This is because in policy experiments, your lever is the policy itself and people can take it up or not as they choose. However, it is still important to try to measure the actual treatment effect on those who took up the program. This will help readers understand if they should, for example, spend resources to encourage more people to sign up for the program, make the program mandatory, or otherwise change the criteria

for entering the program. To do this you will typically utilize the Local Average Treatment Effect (LATE) otherwise known as the Complier Average Causal Effect (CACE).

In this section you should:

- Explain what these two different measurements mean from the perspective of a policymaker
- Which one you believe the policymaker should pay more attention to

Statistical Analysis

Explain the statistical analyses you will actually perform in this section. For a randomized trial, we recommend a simple fixed effects regression, with one fixed effect for any experimental block. Note that the most common type of block is "time of randomization," which commonly occurs when people are rolling into treatment.

In the case of binary outcomes, you may prefer a Cochran-Mantel-Haenszel (CMH) test of 2 x 2 x K tables, where K is the number of blocks. What follows is some boiler plate to this point:

For each outcome (see [meaningful effects](#)), our primary estimate will be the average treatment effect, which is the difference in statistical mean between the measures of treatment and control. We will estimate this difference using block-size weighting using the difference_in_means() command in R (which uses the homoskedastic (HC2)/randomization-justified standard error by default). We will calculate unadjusted 95% confidence intervals for each difference and display them in a plot.

In the case of binary outcomes, our primary test of the null hypothesis of no effects will be the Cochran-Mantel-Haenszel (CMH) test for $2 \times 2 \times K$ tables, where K represents each week in the study period.

In the case of continuous outcomes, our primary test of the null hypothesis of no effects will be a regression $Y_i \sim \alpha + \beta Z_i + \gamma_{b_i} + \varepsilon$, where:

- Y_i is the outcome of participant i,
- α is a constant intercept term,
- β is the average treatment effect of interest,
- Z_i is a binary indicator of whether participant i is in the treatment ($Z_i = 1$) or control ($Z_i = 0$) group,
- γ_{b_i} is a fixed effect for the week (or block) b_i participant i is in,
- ε is random noise, Gaussian in the case of continuous outcomes (annual earnings and pay to hour ratio) and Poisson in the case of count outcomes (number of weeks on unemployment).

Multiple hypothesis adjustment

Commonly, you will have multiple hypotheses you are testing, so you should specify which multiple hypothesis adjustment you will use. Typically, this amounts to explaining the following:

- Do you wish to control the False Detection Rate (FDR) or the Family-wise Error Rate (FWER)?
- Why one over the other?
- Which adjustment will you use (e.g., Bonferroni-Holm for FWER or Benjamini-Hoechberg for FDR)? It is best to specify the code or package that you will use to do this as well.

Balance Test

In this section, you should explain how you will make sure that your experiment didn't go awry. This is especially important for rolling experiments where you're relying on a partner to actually follow the experimental protocol. We recommend at least the following paragraph:

In order to verify that our randomization did not fail, we shall also perform a balance test. Specifically, we will employ the omnibus xBalance test of Hansen and Bowers's RItools package (version 0.1-17) (Hansen & Bowers, 2008). If this omnibus test reports a significant finding (defined for our purposes as a p-value of less than 0.05), then downstream regressions should be viewed skeptically.

C.3 Power Calculations

In this section, you should perform a power analysis specifically asking, for each hypothesis you are testing, what size effect your experiment is powered to detect. These minimum detectable effects should be at least as large as the "meaningful effects" thresholds described in the next section.

Note that power calculations typically require making various assumptions, e.g., about the intra-cluster correlation or about the variance of various effect sizes. If possible it is best to pull representative sizes from the literature review referenced in previous sections.

D. Meaningful Effects

Whenever possible, you should work with agency partners to determine what size of an effect they would consider *meaningful*, by which we mean, what size of effect would cause them to alter their beliefs or modify the intervention under examination. Exactly how you do this is up to you, but some examples might include:

- Having a conversation with stakeholders and asking them what they would consider a meaningful effect;
- A prospective cost-benefit analysis, e.g., "If this messaging strategy gets 1% more people to sign up for health insurance, we will get X more people to sign up, earning $Y more in revenue, and defraying $Z in long-term medical costs."

E. Limitations

Anticipate arguments about why your study doesn't answer the question you say it does or why it doesn't answer "the right" question. For example, perhaps an intervention consists of three components, and people like to talk about one of the three components. You can't make causal claims about the effects of that particular component, and so you should note that as a limitation.

Another common limitation arises in who is *excluded from* the study. Often programs require (non-)participation by certain demographic groups (veterans and minors are common).

THE MISSION OF THE LEGISLATIVE AUDITOR GENERAL IS TO

AUDIT · LEAD · ACHIEVE

WE HELP ORGANIZATIONS IMPROVE