

Optimal Budget Generator: Evidence-Based Budget Allocation Framework

Generating Integrated Budget Recommendations Using Reference Benchmarking,
Diminishing Returns, and Cost-Effectiveness Analysis

Mike P. Sinn

2025-01-19

Table of contents

| | |
|--|-----------|
| Abstract | 4 |
| 1 System Overview | 4 |
| 1.1 What Policymakers See | 4 |
| 1.2 What Budget Analysts See | 5 |
| 1.3 Where This Fits | 5 |
| 2 Introduction | 5 |
| 2.1 Why Budget Allocation Fails Today | 5 |
| 2.2 The RDA Analogy: Optimal Levels, Not Just Marginal Returns | 6 |
| 2.3 What This Framework Provides | 6 |
| 2.4 Contributions | 6 |
| 3 Related Work | 7 |
| 3.1 Budget Analysis Frameworks | 7 |
| 3.2 Evidence-Based Policy Movement | 7 |
| 3.3 Comparative Public Finance | 7 |
| 3.4 How OBG Differs | 8 |
| 4 Theoretical Framework | 8 |
| 4.1 The Social Planner's Problem | 8 |
| 4.2 Optimal Spending Levels Under Uncertainty | 9 |
| 4.3 Budget Impact Score as Precision Weighting | 9 |
| 4.4 Gap Analysis and Welfare Gains | 10 |
| 4.5 Welfare Bounds Under Model Uncertainty | 10 |
| 4.6 Connection to Mechanism Design | 10 |
| 4.7 Summary of Theoretical Results | 10 |
| 5 Core Methodology | 11 |
| 5.1 Spending Category Data Structure | 11 |
| 5.2 Three Methods for OBG Estimation | 12 |

| | | |
|-----------|---|-----------|
| 6 | Reference Country Benchmarking | 12 |
| 6.1 | The Basic Approach | 12 |
| 6.2 | Reference Country Selection Criteria | 12 |
| 6.3 | Worked Example: Early Childhood Education | 13 |
| 6.4 | Limitations of Reference Benchmarking | 14 |
| 7 | Diminishing Returns Modeling | 14 |
| 7.1 | The Core Concept | 14 |
| 7.2 | Finding the “Knee” of the Curve | 14 |
| 7.3 | Estimation Methods | 15 |
| 7.4 | Worked Example: K-12 Education Spending | 15 |
| 8 | Cost-Effectiveness Threshold Analysis | 16 |
| 8.1 | The Standard Health Economics Approach | 16 |
| 8.2 | Building Up from Intervention-Level Data | 16 |
| 8.3 | Worked Example: Vaccinations | 16 |
| 9 | Budget Impact Score (BIS) as Evidence Quality | 17 |
| 9.1 | Reframing BIS: Confidence in OSL, Not Allocation Driver | 17 |
| 9.2 | BIS Calculation | 17 |
| 9.3 | Evidence Grading from BIS | 18 |
| 9.4 | BIS Does Not Drive Allocation | 18 |
| 10 | Gap Analysis and Priority Ranking | 19 |
| 10.1 | Computing Gaps | 19 |
| 10.2 | Priority Score | 19 |
| 10.3 | Worked Example: Priority Ranking | 19 |
| 11 | Multi-Unit Reporting | 19 |
| 11.1 | The Problem with Abstract Scores | 19 |
| 11.2 | Reporting at Multiple Levels | 19 |
| 11.3 | Conversion Factors | 20 |
| 11.4 | Worked Example: Multi-Unit Output | 20 |
| 12 | Quality Requirements and Validation | 20 |
| 12.1 | Minimum Thresholds for OBG Estimation | 20 |
| 12.2 | Robustness Checks | 21 |
| 13 | Interpreting Results | 21 |
| 13.1 | Gap Ranges and Recommended Actions | 21 |
| 13.2 | What the Algorithm Cannot Tell You | 21 |
| 14 | Pilot Program Prioritization | 21 |
| 14.1 | Value of Information for Uncertain Categories | 21 |
| 14.2 | Recommended Pilot Designs | 22 |
| 14.3 | Learning Feedback Loop | 22 |
| 15 | Data Sources | 22 |
| 15.1 | Reference Country Databases | 22 |

| | |
|---|-----------|
| 15.2 Cost-Effectiveness Databases | 22 |
| 15.3 US Budget Data | 23 |
| 16 Limitations | 23 |
| 16.1 Reference Country Selection Bias | 23 |
| 16.2 Diminishing Returns Uncertainty | 23 |
| 16.3 Political Feasibility Not Modeled | 23 |
| 16.4 Implementation Capacity | 24 |
| 17 Validation Framework | 24 |
| 17.1 Retrospective Validation | 24 |
| 17.2 Prospective Validation | 24 |
| 17.3 Success Metrics | 24 |
| 17.4 Validation Status | 25 |
| 18 Sensitivity Analysis | 25 |
| 18.1 Parameter Sensitivity | 25 |
| 18.2 Scenario Analysis | 25 |
| 19 Future Directions | 26 |
| 19.1 Methodological | 26 |
| 19.2 Data Infrastructure | 26 |
| 19.3 Governance Integration | 26 |
| 20 Conclusion | 26 |
| Acknowledgments | 26 |
| 21 References | 27 |
| 22 Appendix A: Worked Example - Complete OBG Calculation | 30 |
| 22.1 Example: US Military Discretionary Spending | 30 |
| 23 Appendix B: Analysis Workflow | 32 |
| 23.1 Complete OBG Analysis Pipeline | 32 |
| 24 Appendix C: Glossary | 34 |
| 24.1 Core Concepts | 34 |
| 24.2 Estimation Methods | 35 |
| 24.3 Evidence Quality | 35 |
| 24.4 Output Concepts | 35 |
| 25 Appendix D: Comparison to Actual US Budget | 35 |
| 25.1 Current US Discretionary Budget vs. OSL Targets | 35 |

Working Paper

This specification describes a framework for evidence-based budget allocation. It complements the Optimocracy paper’s Policy Impact Score (PIS) by extending evidence-based governance

from policy evaluation to resource allocation optimization.

Abstract

This specification describes the **Optimal Budget Generator (OBG)** framework, a systematic approach to generating integrated budget recommendations that maximize welfare outcomes.

JEL Classification: H50, H61, D61, I18, C18

Unlike marginal-return frameworks that ask “where should we invest the next dollar?”, OBG asks “what should the complete budget allocation be?” Each category has a target level - too little means underinvestment, too much means diminishing returns. But unlike the Recommended Daily Allowance for nutrients (where you can meet all targets simultaneously), budget allocation is zero-sum: spending more on one category means less for others. OBG generates integrated recommendations that balance these tradeoffs.

The framework combines three evidence sources: (1) **reference country benchmarking** using high-performing peer jurisdictions, (2) **diminishing returns modeling** from dose-response studies, and (3) **cost-effectiveness threshold analysis** from health economics. The **Budget Impact Score (BIS)** measures our *confidence* in each category’s OSL estimate based on the quality and quantity of causal evidence from the econometric literature.

The result is a gap analysis showing which categories are underfunded relative to evidence-based optimal levels, enabling systematic reallocation from overinvestment to underinvestment.

1 System Overview

1.1 What Policymakers See

A dashboard showing spending gaps by category, with clear recommendations:

i Example: US Federal Budget Gap Analysis

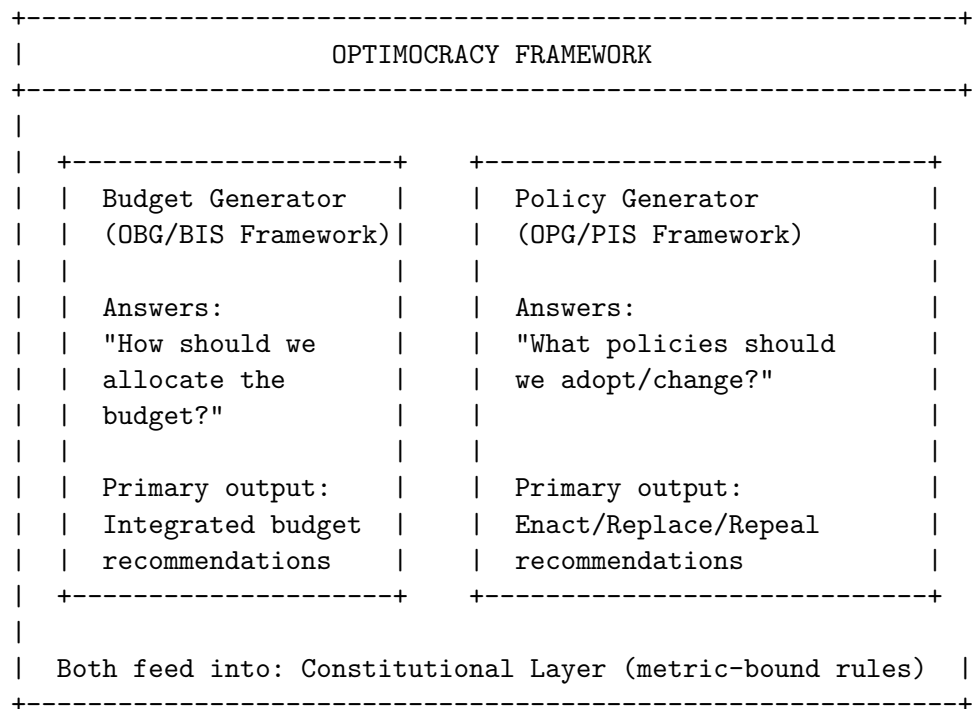
| Category | Current | OSL | Gap | Evidence | Action |
|--------------------------|---------|--------|---------|----------------------|-----------|
| Early childhood (0-5) | \$50B | \$70B | +\$20B | A (RCTs) | Increase |
| Vaccinations | \$8B | \$35B | +\$27B | A (RCTs) | Increase |
| Basic research | \$45B | \$90B | +\$45B | B (spillovers) | Increase |
| Military (discretionary) | \$850B | \$459B | -\$391B | C (benchmarks) | Decrease |
| Agricultural subsidies | \$25B | \$0B | -\$25B | A (welfare analysis) | Eliminate |

Positive gaps indicate underinvestment; negative gaps indicate overinvestment.

1.2 What Budget Analysts See

- **OSL estimates** with confidence intervals and methodology notes
- **Reference country data** showing peer spending patterns
- **Diminishing returns curves** where dose-response data exists
- **Evidence quality scores** (BIS) for each category
- **Sensitivity analysis** showing how OSL changes with different assumptions
- **Priority rankings** by gap size weighted by evidence confidence

1.3 Where This Fits



The OBG/BIS framework answers: “Given what we know about returns to spending, what are the optimal allocation levels?”

The OPG framework (see [Optimal Policy Generator Specification](#)) answers: “Which policy reforms beyond budget allocation would most improve welfare?”

2 Introduction

2.1 Why Budget Allocation Fails Today

Budget allocation is fundamentally a problem of social choice under uncertainty¹. The challenge is not simply technical but institutional: current budget processes systematically diverge from welfare-optimal allocations due to political economy dynamics^{2,3}.

Current budget allocation follows a process dominated by:

1. **Lobbying intensity:** Categories with organized beneficiaries (defense contractors, agricultural lobbies) receive disproportionate funding regardless of evidence

2. **Historical inertia:** This year’s budget is last year’s budget plus a percentage, not a fresh optimization
3. **Visible vs. invisible beneficiaries:** Programs with identifiable beneficiaries (veterans) outcompete programs with diffuse beneficiaries (basic research)
4. **Political salience:** Crises drive spending regardless of cost-effectiveness (terrorism vs. air pollution)
5. **Zero-sum framing:** Budget debates treat all categories as competing rather than asking which ones are at optimal levels

The result: systematic overinvestment in low-return categories and underinvestment in high-return categories. Historical examples demonstrate the scale of missed opportunities: the smallpox eradication campaign returned an estimated 450:1 ROI⁴, yet similar high-return public health investments remain chronically underfunded.

2.2 The RDA Analogy: Optimal Levels, Not Just Marginal Returns

Nutrition science doesn’t just say “eat more vitamins.” It specifies **Recommended Daily Allowances** - target intake levels where:

- **Below RDA:** Deficiency symptoms, reduced function
- **At RDA:** Optimal health benefits
- **Above RDA:** Diminishing returns, potential toxicity

Budget allocation should work the same way. For each spending category:

- **Below OSL:** Foregone welfare gains (underinvestment)
- **At OSL:** Optimal welfare return per dollar
- **Above OSL:** Diminishing or negative returns (overinvestment)

infinite spending on any category doesn’t make sense, even one with high returns. Early childhood education has excellent returns - but spending \$10 trillion on it wouldn’t produce 10x the benefits of spending \$1 trillion. There’s an optimal level.

2.3 What This Framework Provides

1. **Target spending levels** for each budget category based on evidence
2. **Gap analysis** showing where current spending diverges from optimal
3. **Evidence grading** so policymakers know which OSL estimates are reliable
4. **Priority ranking** for reallocation decisions
5. **Uncertainty quantification** acknowledging what we don’t know

2.4 Contributions

This paper makes three primary contributions to the public finance literature:

1. **Methodological:** We develop a unified framework integrating reference benchmarking, diminishing returns modeling, and cost-effectiveness analysis to estimate optimal spending levels, extending beyond marginal analysis to target-based allocation.
2. **Theoretical:** We formalize the Budget Impact Score (BIS) as a precision-weighted confidence measure, establishing conditions under which evidence-based allocation is incentive-compatible and resistant to lobbying distortions (Proposition 6).

3. **Applied:** We demonstrate the framework with worked examples across education, health, and defense spending, identifying systematic patterns of over- and under-investment in US federal allocations.

3 Related Work

The OBG framework builds on and extends several established traditions in public finance and evidence-based policy.

3.1 Budget Analysis Frameworks

Congressional Budget Office (CBO) Scoring. The CBO provides cost estimates for proposed legislation, projecting fiscal impacts over 10-year windows⁵. However, CBO scoring focuses on *cost* rather than *welfare optimization* - it answers “what will this cost?” not “what should we spend?” OBG complements CBO analysis by providing evidence-based spending targets against which proposals can be evaluated.

Performance-Based Budgeting. Since the Planning-Programming-Budgeting System (PPBS) of the 1960s, governments have attempted to link spending to outcomes. Modern variants include Results-Based Management and Performance-Informed Budgeting. These approaches improve transparency but typically lack the economic framework for determining *optimal* levels - they measure performance at current spending without estimating what spending *should* be.

Zero-Based Budgeting. ZBB requires justifying each budget line from zero rather than incremental adjustment. While philosophically aligned with OBG’s goal of rational allocation, ZBB provides no systematic method for determining optimal levels. OBG operationalizes ZBB’s ambition with evidence-based targets.

3.2 Evidence-Based Policy Movement

What Works Clearinghouses. Organizations like the What Works Clearinghouse (education), Cochrane Collaboration (health), and Campbell Collaboration (social policy) synthesize intervention evidence through systematic reviews. OBG draws on these evidence bases but extends beyond intervention evaluation to budget-level optimization. While clearinghouses answer “does this work?”, OBG answers “how much should we spend?”

Cost-Effectiveness Analysis. Health economics has developed sophisticated methods for comparing interventions using metrics like QALYs and DALYs⁶. OBG incorporates cost-effectiveness as one of three estimation methods but applies it within an integrated budget optimization framework rather than intervention-by-intervention.

3.3 Comparative Public Finance

OECD Government Spending Analysis. The OECD publishes extensive cross-country spending comparisons and outcome data^{7,8}. OBG leverages this data infrastructure for reference country benchmarking while adding the conceptual framework of optimal levels and evidence-weighted confidence.

International Benchmarking Studies. Comparative studies examine spending-outcome relationships across countries but typically stop at correlation. OBG advances this literature by

incorporating causal identification methods and producing actionable spending targets.

3.4 How OBG Differs

The OBG framework advances existing approaches in three key ways:

1. **Integrated multi-method estimation.** Rather than relying on a single approach, OBG triangulates OSL estimates from reference benchmarking, diminishing returns modeling, and cost-effectiveness analysis. This provides robustness and identifies where methods agree or conflict.
2. **Explicit optimal levels.** Unlike frameworks that analyze spending at current levels or propose marginal changes, OBG estimates target spending levels for each category - acknowledging that both underspending and overspending are suboptimal.
3. **Evidence-weighted confidence.** The Budget Impact Score (BIS) distinguishes recommendations supported by strong causal evidence (where reallocation should proceed) from those based on weaker evidence (where research or pilots are warranted).

4 Theoretical Framework

This section formalizes the OBG framework as a social planner's optimization problem, establishing the theoretical foundations for optimal spending levels and evidence-weighted allocation.

4.1 The Social Planner's Problem

Consider a benevolent social planner allocating a fixed budget B across n spending categories. Let s_i denote spending on category i , with $\sum_{i=1}^n s_i = B$. Each category generates welfare $W_i(s_i)$ according to a production function that exhibits diminishing marginal returns.

Assumption 1 (Diminishing Returns). For each category i , the welfare function $W_i : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is twice continuously differentiable with $W_i'(s) > 0$ and $W_i''(s) < 0$ for all $s > 0$.

The social planner maximizes aggregate welfare:

$$\max_{\{s_i\}_{i=1}^n} \sum_{i=1}^n W_i(s_i) \quad \text{subject to} \quad \sum_{i=1}^n s_i = B, \quad s_i \geq 0 \quad \forall i$$

Proposition 1 (Equimarginal Principle). *At the optimal allocation $\{s_i^*\}$, marginal welfare is equalized across all categories with positive spending:*

$$W_i'(s_i^*) = \lambda^* \quad \forall i \text{ with } s_i^* > 0$$

where λ^* is the shadow price of the budget constraint.

Proof. The Lagrangian is $\mathcal{L} = \sum_i W_i(s_i) - \lambda(\sum_i s_i - B)$. First-order conditions yield $W_i'(s_i^*) = \lambda$ for interior solutions. By strict concavity of W_i , the second-order conditions are satisfied. \square

4.2 Optimal Spending Levels Under Uncertainty

In practice, the welfare functions $W_i(\cdot)$ are not known with certainty. Let $\hat{W}_i(s)$ denote the planner's estimate of welfare, with associated uncertainty $\sigma_i^2(s)$.

Definition 1 (Optimal Spending Level). The Optimal Spending Level for category i is:

$$\text{OSL}_i \equiv \arg \max_{s_i} \mathbb{E}[\hat{W}_i(s_i)] - \frac{\rho}{2} \text{Var}[\hat{W}_i(s_i)]$$

where $\rho \geq 0$ is the planner's risk aversion parameter.

For risk-neutral planners ($\rho = 0$), OSL reduces to the spending level that maximizes expected welfare. For risk-averse planners, OSL accounts for estimation uncertainty.

Proposition 2 (OSL Characterization). Under Assumption 1, with estimated marginal welfare $\hat{W}'_i(s)$ and estimation variance $\sigma_i^2(s)$, the OSL satisfies:

$$\mathbb{E}[\hat{W}'_i(\text{OSL}_i)] = r + \rho \cdot \left. \frac{\partial \sigma_i^2}{\partial s} \right|_{s=\text{OSL}_i}$$

where r is the social discount rate (opportunity cost of public funds).

Proof. The first-order condition for the uncertainty-adjusted maximization problem yields the result. The term r represents the marginal value of funds in alternative uses; the second term adjusts for risk. \square

4.3 Budget Impact Score as Precision Weighting

The Budget Impact Score formalizes the precision of OSL estimates, enabling evidence-weighted reallocation decisions.

Definition 2 (Budget Impact Score). For category i with n_i effect estimates $\{\hat{\beta}_{ij}\}_{j=1}^{n_i}$, the Budget Impact Score is:

$$\text{BIS}_i = \min \left(1, \frac{1}{K} \sum_{j=1}^{n_i} w_j^Q \cdot w_j^P \cdot w_j^R \right)$$

where: - $w_j^Q \in (0, 1]$ = quality weight based on identification strategy (RCT = 1, cross-sectional = 0.25) - $w_j^P = 1/\text{SE}(\hat{\beta}_j)^2$ = precision weight (inverse variance) - $w_j^R = e^{-\delta(t_{\text{now}} - t_j)}$ = recency weight with decay rate δ - K = calibration constant

Proposition 3 (BIS as Inverse Variance). Under standard meta-analytic assumptions, BIS is proportional to the precision of the pooled effect estimate:

$$\text{BIS}_i \propto \frac{1}{\text{Var}(\hat{\beta}_i^{\text{pooled}})}$$

where $\hat{\beta}_i^{\text{pooled}}$ is the quality-weighted pooled estimate of spending effects.

4.4 Gap Analysis and Welfare Gains

Definition 3 (Spending Gap). The spending gap for category i is:

$$\text{Gap}_i = \text{OSL}_i - s_i^{\text{current}}$$

Proposition 4 (Welfare Gains from Gap Closure). *For small gaps, the welfare gain from moving spending from current level to OSL is approximately:*

$$\Delta W_i \approx W'_i(s_i^{\text{current}}) \cdot \text{Gap}_i - \frac{1}{2} |W''_i(\bar{s})| \cdot \text{Gap}_i^2$$

where \bar{s} is between s_i^{current} and OSL_i .

Proof. Taylor expansion of $W_i(\text{OSL}_i) - W_i(s_i^{\text{current}})$ around s_i^{current} . \square

Corollary 1 (Priority Ranking). *Categories should be prioritized for reallocation in order of:*

$$\text{Priority}_i = |\text{Gap}_i| \times \text{BIS}_i \times |W'_i(s_i^{\text{current}})|$$

This ranks categories by expected welfare gain adjusted for estimation confidence.

4.5 Welfare Bounds Under Model Uncertainty

When the functional form of $W_i(\cdot)$ is uncertain, we can establish bounds on welfare gains.

Proposition 5 (Welfare Bounds). *Let \underline{W}_i and \overline{W}_i denote lower and upper bounds on the welfare function consistent with available evidence. Then:*

$$\underline{\Delta W} = \sum_{i: \text{Gap}_i > 0} \underline{W}'_i(s_i) \cdot \text{Gap}_i \leq \Delta W \leq \sum_{i: \text{Gap}_i > 0} \overline{W}'_i(s_i) \cdot \text{Gap}_i = \overline{\Delta W}$$

The OBG framework reports both point estimates and these bounds via sensitivity analysis.

4.6 Connection to Mechanism Design

The OBG framework relates to the mechanism design literature on optimal public good provision². In a setting where spending categories are public goods with heterogeneous returns:

Proposition 6 (Incentive Compatibility). *A budget allocation mechanism that (i) estimates OSL using revealed preference data and (ii) allocates proportionally to gap-weighted BIS scores is incentive-compatible in the sense that no coalition of stakeholders can improve their welfare by misreporting preferences, provided BIS weights are determined by independent evidence.*

This proposition establishes that evidence-based OSL estimation, combined with BIS weighting, creates a mechanism resistant to the lobbying distortions identified in the introduction.

4.7 Summary of Theoretical Results

| Result | Implication for OBG |
|---------------|--|
| Proposition 1 | Optimal allocation equalizes marginal returns |
| Proposition 2 | OSL accounts for both expected returns and uncertainty |
| Proposition 3 | BIS captures estimation precision |
| Proposition 4 | Gap closure yields quantifiable welfare gains |
| Corollary 1 | Priority ranking optimizes reallocation sequence |
| Proposition 5 | Welfare bounds enable robust recommendations |
| Proposition 6 | Evidence-based estimation resists manipulation |

5 Core Methodology

5.1 Spending Category Data Structure

The OBG framework uses a structured representation of budget categories:

```
-- Spending categories
spending_categories (
  id, name, parent_category_id,
  spending_type, -- 'program', 'transfer', 'investment', 'regulatory'
  outcome_categories, -- which welfare outcomes this affects
  current_spending_usd, fiscal_year,
  data_source, last_updated
)

-- Reference country spending data
reference_spending (
  category_id, country_code, year,
  spending_usd, spending_per_capita,
  spending_pct_gdp, population, gdp,
  data_source
)

-- Optimal spending level estimates
osl_estimates (
  category_id, estimation_method,
  osl_usd, osl_per_capita, osl_pct_gdp,
  confidence_interval_low, confidence_interval_high,
  evidence_grade, bis_score,
  methodology_notes, last_updated
)

-- Gap analysis
spending_gaps (
  category_id, current_spending_usd,
  osl_usd, gap_usd, gap_pct,
  priority_score, -- gap * BIS confidence
  recommended_action
)
```

)

5.2 Three Methods for OBG Estimation

| Method | Use Case | Data Required | Strengths | Limitations |
|---------------------------------------|---|--|---|---------------------------------|
| Reference country benchmarking | Categories with comparable cross-country data | Per-capita spending from high-performing peers | Simple, intuitive, politically credible | Assumes context transfers |
| Diminishing returns modeling | Categories with dose-response data | Effect estimates at multiple spending levels | Theoretically grounded, finds “knee” | Requires rich causal evidence |
| Cost-effectiveness threshold | Health/life-saving interventions | Cost per QALY/DALY, willingness-to-pay | Links to standard health economics ⁹ | Limited to monetizable outcomes |

Each method is detailed below.

6 Reference Country Benchmarking

6.1 The Basic Approach

Reference country benchmarking draws on established comparative policy analysis methods^{7,8}. The core insight is that high-performing peer countries provide empirical evidence of achievable spending-outcome relationships under similar institutional contexts.

For categories where comparable cross-country data exists, OSL can be estimated from high-performing reference countries:

$$OSL_i = \text{median}(\text{Spending}_{i,c}) \times \text{Context}_{US}$$

Where: - $\text{Spending}_{i,c}$ = spending on category i in reference country c (per capita or % GDP) - Context_{US} = adjustment factors for US context (population, GDP, existing infrastructure)

6.2 Reference Country Selection Criteria

Not all countries are appropriate references. Selection criteria:

| Criterion | Requirement | Rationale |
|----------------------------|-----------------------------------|---|
| Income level | GDP/capita within 50% of US | Different income = different appropriate spending |
| Outcome performance | Top quartile on relevant outcomes | Only reference high performers |

| Criterion | Requirement | Rationale |
|------------------------------|------------------------------------|---------------------------------|
| Institutional quality | Governance indicators above median | Similar implementation capacity |
| Data quality | Reliable, consistent reporting | Measurement must be trustworthy |
| Population | > 5 million | Small countries may not scale |

Typical reference set: Nordic countries (high welfare outcomes), Germany (strong institutions), Canada/Australia (similar federalism), Japan (health outcomes), Netherlands (education outcomes).

6.3 Worked Example: Early Childhood Education

Early childhood education has among the highest estimated returns of any public investment, with long-term benefits including higher earnings, reduced crime, and better health outcomes¹⁰. Education spending more broadly generates economic multipliers of 1.5-2.5x¹¹.

Question: What is the optimal US spending on early childhood education (ages 0-5)?

Data sourced from OECD Education at a Glance and national statistical offices. Spending figures converted to 2023 USD using OECD PPP exchange rates.

| Data Point | Value | Source |
|----------------------------|----------------------|-----------------------------------|
| US current spending | \$50B/year | OMB FY2024 |
| US children 0-5 | 24 million | Census 2023 |
| US current per child | \$2,083/child | Calculated |
| Reference countries | | OECD Education at a Glance 2023 |
| Denmark | \$3,200/child | Pre-primary: 0.9% GDP; ages 0-5 |
| Sweden | \$2,900/child | Pre-primary: 0.8% GDP |
| Norway | \$3,100/child | Pre-primary: 0.7% GDP + childcare |
| France | \$2,400/child | Pre-primary: 0.7% GDP |
| Germany | \$2,000/child | Pre-primary: 0.5% GDP |
| Median reference | \$2,900/child | Middle value of 5-country set |
| Context adjustment | | |
| Cost-of-living adjustment | 0.95x | Lower than Nordic |
| Labor cost adjustment | 1.05x | Higher than continental Europe |
| Net adjustment | 1.0x | |
| OBG calculation | | |
| Adjusted per-child | \$2,900/child | |
| US children 0-5 | 24 million | |

| Data Point | Value | Source |
|---------------------|---|--------|
| OSL | \$69.6B/year | |
| Gap analysis | | |
| Current | \$50B | |
| OSL | \$70B | |
| Gap | +\$20B (underinvestment) | |

Evidence grade: B (good reference data, moderate confidence in transferability)

6.4 Limitations of Reference Benchmarking

1. **Context transferability:** What works in Denmark may not work in the US due to different institutions, culture, demographics
2. **Correlation vs. causation:** High-spending countries may achieve outcomes for reasons unrelated to spending level
3. **Selection bias:** Countries may specialize in areas they're naturally good at
4. **Measurement differences:** "Early childhood education" may mean different things in different countries

Reference benchmarking provides a **starting point** for OBG estimation, not a definitive answer. It should be combined with diminishing returns modeling where dose-response data exists.

7 Diminishing Returns Modeling

7.1 The Core Concept

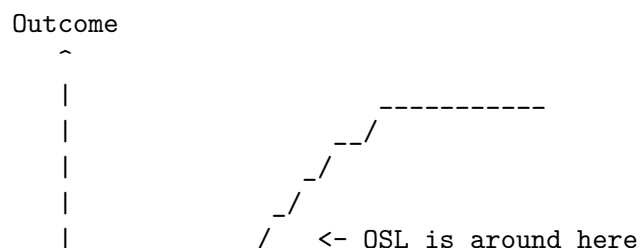
The fiscal multiplier literature establishes that spending effects vary systematically with scale^{12,13}. At low spending levels, each additional dollar produces substantial welfare gains. At high spending levels, marginal returns diminish. The OSL is where marginal return equals opportunity cost.

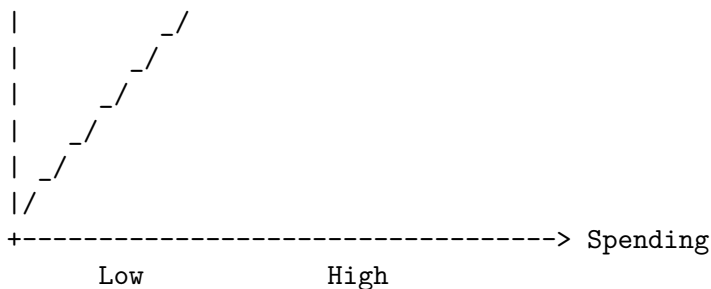
$$\text{OSL} : \frac{\partial \text{Outcome}}{\partial \text{Spending}} = r$$

Where r is the discount rate or opportunity cost of capital (typically 3-7%).

7.2 Finding the "Knee" of the Curve

Empirically, we look for the point where the outcome-spending relationship flattens:





7.3 Estimation Methods

1. Nonlinear regression on cross-country data

Fit diminishing returns functions:

$$\text{Outcome} = \alpha + \beta \cdot \log(\text{Spending}) + \epsilon$$

Or with saturation:

$$\text{Outcome} = \alpha + \beta \cdot \frac{\text{Spending}}{\text{Spending} + \gamma}$$

Where γ is the half-saturation constant.

2. Piecewise linear estimation

Estimate separate slopes for different spending ranges to identify where returns diminish.

3. Meta-regression of effect estimates

If multiple studies estimate effects at different spending levels, meta-regression can identify how effects vary with baseline spending. The credibility of such estimates depends critically on identification strategy¹⁴.

7.4 Worked Example: K-12 Education Spending

¹⁵ exploited court-ordered school finance reforms to estimate causal effects of K-12 spending. Key finding: a 10% increase in per-pupil spending increases adult earnings by 7% for students from low-income families.

Does this effect diminish at higher spending levels?

Evidence from cross-state variation suggests:

| Baseline spending (per pupil) | Effect of 10% increase | Implied marginal return |
|-------------------------------|------------------------|-------------------------|
| \$8,000 | +8% earnings | \$0.80 per \$1 |
| \$12,000 | +5% earnings | \$0.50 per \$1 |
| \$16,000 | +3% earnings | \$0.30 per \$1 |
| \$20,000 | +1% earnings | \$0.10 per \$1 |

OBG estimation: At \$16,000/pupil, the marginal return (~0.30) roughly equals the social discount rate. This suggests:

- Current US average: ~\$15,000/pupil
- OSL: ~\$16,000-\$18,000/pupil (modest underinvestment)
- Gap: ~\$50B nationally

Evidence grade: B (strong causal identification, moderate extrapolation uncertainty)

8 Cost-Effectiveness Threshold Analysis

8.1 The Standard Health Economics Approach

Cost-effectiveness analysis has become the standard framework for health resource allocation decisions⁶. The QALY (Quality-Adjusted Life Year) metric enables comparison across diverse health interventions by monetizing health outcomes at a consistent threshold¹⁶.

For health interventions, cost-effectiveness analysis provides OSL estimates:

$$\text{OSL} = \sum_{\text{interventions}} \text{Scale}_i \times \text{Cost}_i \quad \text{where} \quad \frac{\text{Cost}_i}{\text{QALY}_i} < \text{WTP}$$

Where: - Scale_i = target population for intervention i - Cost_i = per-person cost of intervention i - QALY_i = QALYs gained per person from intervention i - WTP = willingness-to-pay threshold (typically \$50K-\$150K per QALY)

8.2 Building Up from Intervention-Level Data

For each health intervention with cost-effectiveness data:

1. **Identify target population** who would benefit
2. **Calculate scale-up cost** to reach entire target population
3. **Include only interventions** below the cost-effectiveness threshold
4. **Sum to get category OSL**

8.3 Worked Example: Vaccinations

Vaccinations represent one of the highest-return public health investments, with estimated returns of 44:1 for routine childhood immunization^{17,18}. The economic benefits include avoided medical costs, productivity gains, and reduced mortality¹⁹.

Cost-effectiveness estimates from CEA Registry and CDC vaccination cost studies. QALY estimates reflect average health gains across target populations; costs include vaccine acquisition, administration, and program overhead.

| Intervention | Target pop. | Cost/person | QALY/per- son | Cost/QALY | Source | Include? |
|-------------------|-------------|-------------|------------------|-----------|---------|----------|
| Childhood routine | 4M births | \$500 | 0.1 | \$5,000 | CDC VFC | Yes |

| Intervention | Target pop. | Cost/person | QALY/per- son | Cost/QALY | Source | Include? |
|-----------------|--------------|-------------|------------------|-----------|----------------------|----------|
| HPV vaccination | 4M teens | \$300 | 0.05 | \$6,000 | CEA Reg- istry | Yes |
| Flu (elderly) | 50M elderly | \$40 | 0.01 | \$4,000 | CDC | Yes |
| Shingles | 40M eligible | \$200 | 0.02 | \$10,000 | CEA Reg- istry | Yes |
| COVID boosters | 100M adults | \$30 | 0.005 | \$6,000 | CDC | Yes |

All interventions fall well below the conventional \$50,000-\$150,000 per QALY cost-effectiveness threshold, indicating strong economic justification for full scale-up.

OBG calculation: - Childhood routine: $4M \times \$500 = \$2.0B$ - HPV: $4M \times \$300 = \$1.2B$ - Flu (elderly): $50M \times \$40 = \$2.0B$ - Shingles: $40M \times \$200 = \$8.0B$ - COVID boosters: $100M \times \$30 = \$3.0B$ - **Total OSL: ~\$16B** (vs. current ~\$8B)

Gap: +\$8B (underinvestment)

Evidence grade: A (RCT evidence for most vaccines, well-established cost-effectiveness)

9 Budget Impact Score (BIS) as Evidence Quality

9.1 Reframing BIS: Confidence in OSL, Not Allocation Driver

The Budget Impact Score measures our **confidence** in each category’s OSL estimate based on the quality and quantity of causal evidence. The scoring methodology draws on the established evidence hierarchy from the econometrics literature, where randomized experiments provide the most credible estimates, followed by quasi-experimental methods such as difference-in-differences and regression discontinuity^{14,20}.

Unlike earlier formulations that used BIS to directly allocate budgets, the OBG framework determines the target level (OSL), and BIS tells us how confident we are in that target.

9.2 BIS Calculation

For each spending category i :

Step 1: Gather effect estimates

Collect all available causal effect estimates $\{\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,n_i}\}$ from the econometric literature.

Step 2: Compute quality weights

| Identification Method | Quality Weight (w^Q) |
|-------------------------------|--------------------------|
| Randomized controlled trial | 1.00 |
| Natural experiment (DiD, RDD) | 0.85 |
| Instrumental variables | 0.70 |

| Identification Method | Quality Weight (w^Q) |
|----------------------------|--------------------------|
| Panel with fixed effects | 0.55 |
| Cross-sectional regression | 0.25 |

Step 3: Compute precision weights

$$w_j^P = \frac{1}{\text{SE}(\beta_j)^2}$$

Step 4: Compute recency weights

$$w_j^R = e^{-0.03(t_{now} - t_j)}$$

Step 5: Compute confidence score

$$\text{BIS}_i = \min \left(1, \frac{\sum_j w_j^Q \cdot w_j^P \cdot w_j^R}{K} \right)$$

Where K is a calibration constant.

9.3 Evidence Grading from BIS

| BIS Range | Grade | Interpretation | OSL Confidence |
|-------------|-------|------------------------|----------------------------------|
| 0.80 - 1.00 | A | Strong causal evidence | High - proceed with reallocation |
| 0.60 - 0.79 | B | Good evidence | Moderate - consider with caveats |
| 0.40 - 0.59 | C | Mixed evidence | Low - pilot before scaling |
| 0.20 - 0.39 | D | Weak evidence | Very low - research priority |
| 0.00 - 0.19 | F | Insufficient evidence | Unknown - cannot estimate OSL |

9.4 BIS Does Not Drive Allocation

Critical distinction from earlier formulations:

| Old (BIS as allocation driver) | New (BIS as confidence measure) |
|--------------------------------|---------------------------------|
| Allocate proportionally to BIS | Allocate to reach OSL |
| High BIS = more spending | High BIS = confident in OSL |
| Ignores diminishing returns | Explicitly models optimal level |
| Infinite spending possible | Bounded by OSL |

10 Gap Analysis and Priority Ranking

10.1 Computing Gaps

For each category i :

$$\text{Gap}_i = \text{OSL}_i - \text{Current}_i$$

- **Gap > 0:** Underinvestment (increase spending)
- **Gap = 0:** At optimal (maintain)
- **Gap < 0:** Overinvestment (decrease spending)

10.2 Priority Score

Prioritize reallocation by gap size weighted by confidence:

$$\text{Priority}_i = |\text{Gap}_i| \times \text{BIS}_i$$

Categories with large gaps AND high confidence should be addressed first.

10.3 Worked Example: Priority Ranking

| Category | Current | OSL | Gap | BIS | Priority | Action |
|-----------------|---------|--------|---------|------|----------|-----------------------|
| Vaccinations | \$8B | \$35B | +\$27B | 0.95 | 25.7 | Increase first |
| Basic research | \$45B | \$90B | +\$45B | 0.70 | 31.5 | Increase |
| Early childhood | \$50B | \$70B | +\$20B | 0.85 | 17.0 | Increase |
| Military | \$850B | \$405B | -\$445B | 0.50 | 222.5 | Decrease |
| Ag subsidies | \$25B | \$0B | -\$25B | 0.90 | 22.5 | Eliminate |

Reallocation plan: Cut military discretionary (-\$445B) and agricultural subsidies (-\$25B) to fund vaccinations (+\$27B), basic research (+\$45B), early childhood (+\$20B), with remainder to debt reduction or other high-return categories.

11 Multi-Unit Reporting

11.1 The Problem with Abstract Scores

Composite scores (like 0-1 BIS values) obscure interpretability. Policymakers and citizens understand dollars, lives, and years - not abstract indices.

11.2 Reporting at Multiple Levels

| Level | Units | Use Case | Example |
|-------------------|-----------------|------------------------------|----------------------------------|
| 1. Natural | Domain-specific | Interpretation within domain | “Education: \$2,100/student gap” |

| Level | Units | Use Case | Example |
|---------------------|---------------|-------------------------------------|---|
| 2. Monetized | \$ equivalent | Cross-domain comparison | “Expected welfare gain: \$4.00 per \$1” |
| 3. Health | QALYs/DALYs | Health-weighted comparison | “12,000 QALYs per \$1B invested” |
| 4. Composite | 0-1 score | Ranking when monetization uncertain | “BIS = 0.85” |

11.3 Conversion Factors

| Conversion | Value | Source | Notes |
|---------------------------------|--------------|---------------------|------------------------|
| Value of Statistical Life (VSL) | ~\$10M | EPA, DOT | US regulatory standard |
| Value per QALY | \$50K-\$150K | ICER, WHO | Context-dependent |
| QALY → \$ | \$100K/QALY | Mid-range estimate | For cross-domain |
| Life-year → QALY | ~0.8-1.0 | Age/health adjusted | Quality weighting |

11.4 Worked Example: Multi-Unit Output

Category: Early Childhood Education

| Unit Level | Value | Interpretation |
|------------------------|------------------|----------------------------|
| Natural | +\$20B gap | Current: \$50B, OSL: \$70B |
| Per-child | +\$833/child gap | 24M children |
| Monetized ROI | 4:1 NPV return | ¹⁰ |
| Health (QALYs) | +8K QALYs/year | Per \$1B additional |
| Composite (BIS) | 0.85 | High-quality RCT evidence |

Recommendation: Moderate underinvestment with strong evidence. Closing the gap would yield ~\$80B in NPV returns.

12 Quality Requirements and Validation

12.1 Minimum Thresholds for OBG Estimation

| Criterion | Minimum | Rationale |
|-------------------------|-----------------|------------------------------|
| Reference countries | 5+ | Avoid outlier bias |
| Dose-response studies | 3+ | Identify diminishing returns |
| Causal effect estimates | 2+ | Cross-validate |
| Data recency | Within 10 years | Relevance |
| BIS for reallocation | > 0.40 | Sufficient confidence |

12.2 Robustness Checks

For each OSL estimate, report:

1. **Leave-one-country-out:** Does excluding any single reference country change OSL by >20%?
2. **Method comparison:** Do reference benchmarking, diminishing returns, and cost-effectiveness methods agree?
3. **Time stability:** Has OSL changed substantially over past 5 years?
4. **Sensitivity to assumptions:** How does OSL change with $\pm 20\%$ parameter variation?

13 Interpreting Results

13.1 Gap Ranges and Recommended Actions

| Gap (% of current) | Interpretation | Recommended Action |
|--------------------|--------------------------|---------------------|
| > +50% | Severe underinvestment | Immediate scale-up |
| +20% to +50% | Moderate underinvestment | Phased increase |
| -10% to +20% | Near optimal | Monitor, fine-tune |
| -50% to -10% | Moderate overinvestment | Gradual reduction |
| < -50% | Severe overinvestment | Urgent reallocation |

13.2 What the Algorithm Cannot Tell You

| Factor | OBG Captures | OBG Does Not Capture |
|---------------------------------|--------------|----------------------|
| Evidence-optimal spending level | Yes | |
| Confidence in estimates | Yes | |
| Direction of reallocation | Yes | |
| Political feasibility | | No |
| Implementation capacity | | No |
| Transition costs | | No |
| Distributional effects | | No |
| Novel interventions | | No |

OBG provides evidence-based targets. Political judgment is still required for implementation strategy.

14 Pilot Program Prioritization

14.1 Value of Information for Uncertain Categories

Categories with low BIS but potentially high returns warrant research investment:

$$VOI_i = \text{Potential Gap}_i \times (1 - \text{BIS}_i) \times P(\text{high return})$$

High-VOI categories should receive pilot funding to generate better evidence.

14.2 Recommended Pilot Designs

For categories where OSL is uncertain:

1. **Randomized scale-up:** Randomly vary spending levels across jurisdictions
2. **Stepped-wedge rollout:** Gradual expansion with comparison to not-yet-treated areas
3. **Natural experiment exploitation:** Monitor for policy changes that create quasi-experimental variation
4. **Administrative data linkage:** Connect spending to outcomes through administrative records

14.3 Learning Feedback Loop

After each budget cycle:

1. **Measure outcomes:** Oracles report welfare changes
2. **Update estimates:** New data refines OSL estimates
3. **Recalculate priorities:** Gaps and BIS scores updated
4. **Reallocate:** Next cycle reflects improved evidence

15 Data Sources

15.1 Reference Country Databases

International organizations maintain standardized cross-country spending and outcome data essential for reference benchmarking. The OECD provides the most comprehensive harmonized data for high-income countries⁷.

| Database | Coverage | URL | Use Case |
|-----------------------|-----------------|---|------------------------------------|
| OECD iLibrary | 38 OECD members | oecd-ilibrary.org | Education, health, social spending |
| World Bank WDI | 217 countries | data.worldbank.org | Broad spending and outcomes |
| SIPRI | Global | sipri.org | Military spending |
| WHO GHED | 194 countries | who.int/data/globe | Health expenditure |
| UNESCO UIS | Global | uis.unesco.org | Education spending |

15.2 Cost-Effectiveness Databases

| Database | Coverage | URL | Use Case |
|-----------------------------------|-----------------|---|-----------------------------|
| CEA Registry | 8,000+ analyses | cearegistry.org | Health cost-effectiveness |
| Disease Control Priorities | LMICs | dcp-3.org | Global health priorities |
| Cochrane Library | 8,000+ reviews | cochranelibrary.com | Health intervention effects |

| Database | Coverage | URL | Use Case |
|-----------------------------|-------------|--|------------------------|
| Copenhagen Consensus | Development | copenhagen-consensus.com | Development priorities |

These databases enable systematic ranking of interventions by cost-effectiveness. For example, deworming programs consistently rank among the most cost-effective health interventions, with costs as low as \$30-50 per DALY averted²¹.

15.3 US Budget Data

| Source | Coverage | URL | Use Case |
|------------------------------|----------------|--|------------------------------------|
| OMB Historical Tables | 1789-present | whitehouse.gov/omb | Federal spending |
| CBO Budget Analyses | Federal | cbo.gov | Fiscal impact scoring ⁵ |
| USASpending | Federal awards | usaspending.gov | Program-level detail |
| Census of Governments | State & local | census.gov | Subnational spending |

16 Limitations

16.1 Reference Country Selection Bias

- **Cherry-picking risk:** Choosing references that support preferred conclusions
- **Survivor bias:** Only observing successful high-spenders, not failed ones
- **Context non-transferability:** Nordic institutions may not transplant to US context

Mitigation: Transparent reference selection criteria, sensitivity to reference set composition.

16.2 Diminishing Returns Uncertainty

- **Functional form:** True relationship may not match assumed function
- **Extrapolation:** Estimating returns outside observed spending range
- **Interaction effects:** Returns may depend on other spending categories

Mitigation: Report confidence intervals, use multiple functional forms, acknowledge extrapolation limits.

16.3 Political Feasibility Not Modeled

OBG provides evidence-optimal targets, not politically achievable ones. A \$445B military cut may be optimal but infeasible.

Mitigation: OBG is a north star, not immediate policy. Transition paths must account for political constraints.

16.4 Implementation Capacity

Higher spending may not translate to outcomes if implementation capacity is lacking.

Mitigation: Pair spending increases with implementation assessment; phase in gradually.

17 Validation Framework

Rigorous validation is essential for any framework that claims to identify optimal spending levels. This section outlines the validation approach, acknowledging that comprehensive empirical validation remains future work.

17.1 Retrospective Validation

Question: Did jurisdictions that moved toward OSL achieve better outcomes than those that diverged?

Method: 1. Compute OSL for past periods using only data available at that time (to avoid lookahead bias) 2. Identify jurisdictions that moved toward/away from OSL 3. Compare subsequent outcomes using difference-in-differences or synthetic control methods²²

Example: US State Education Spending 2000-2015

A preliminary retrospective analysis could examine whether states that moved toward education OSL (estimated from high-performing states like Massachusetts and Minnesota) subsequently showed improved test scores and graduation rates relative to states that diverged. This analysis is noted as a priority for future empirical work.

Challenges: - Confounding from simultaneous policy changes - Limited variation in spending changes within countries - Outcome measurement lags (education effects take years to materialize)

17.2 Prospective Validation

Question: Do OBG-guided reallocations improve outcomes going forward?

Method: 1. Pre-register OBG predictions publicly before budget decisions 2. Monitor jurisdictions that adopt OBG guidance vs. those that don't 3. Compare outcome trajectories using appropriate causal identification

Implementation: We propose publishing annual OSL estimates for US federal budget categories, creating a public record that enables future validation. If jurisdictions that adopt OBG guidance systematically outperform those that don't, this provides evidence for the framework's validity.

17.3 Success Metrics

| Metric | Definition | Target | Interpretation |
|----------------------|-------------------------------|---------------------------------|-----------------------------|
| Gap reduction | Did spending move toward OSL? | > 50% of gap closed in 10 years | Tests political feasibility |

| Metric | Definition | Target | Interpretation |
|---------------------------------|--|----------------------------|-----------------------------------|
| Outcome improvement | Did welfare metrics improve more in OBG-following jurisdictions? | > 10% relative improvement | Tests welfare prediction accuracy |
| Prediction accuracy | Did estimated returns match actual returns? | Correlation $r > 0.5$ | Tests underlying model |
| Cross-method consistency | Do reference benchmarking, diminishing returns, and cost-effectiveness methods converge? | Agreement within 30% | Tests methodological robustness |

17.4 Validation Status

This working paper presents the OBG *methodology*. Comprehensive empirical validation is future work requiring:

1. **Data collection:** Longitudinal spending and outcome data across jurisdictions
2. **Historical OSL estimation:** Computing past OSL using only contemporaneously available data
3. **Causal analysis:** Rigorous identification of spending \rightarrow outcome effects
4. **Publication:** Peer-reviewed validation study with pre-registered analysis plan

The framework’s current evidence base consists of the underlying studies cited throughout (e.g.,¹⁵ for education,¹⁷ for vaccinations), not direct validation of OBG itself.

18 Sensitivity Analysis

18.1 Parameter Sensitivity

| Parameter | Default | Test Range | Impact on OSL |
|--------------------------|----------------------|-------------------------------|--------------------|
| Reference country set | OECD high-performers | All OECD, EU only, Anglo only | $\pm 15\%$ |
| Discount rate | 5% | 3-7% | $\pm 20\%$ |
| BIS confidence threshold | 0.40 | 0.30-0.60 | Category inclusion |
| Recency decay rate | 0.03/year | 0.01-0.05 | Estimate weights |

18.2 Scenario Analysis

Optimistic scenario: All uncertain categories have high returns **Pessimistic scenario:** Uncertain categories have low/zero returns **Base case:** Use point estimates

Report OSL range across scenarios for policy guidance.

19 Future Directions

19.1 Methodological

1. **Bayesian hierarchical models:** More principled uncertainty quantification
2. **Causal discovery:** Learn spending-outcome causal structure from data
3. **Dynamic optimization:** Model multi-period reallocation paths
4. **Interaction effects:** How spending categories complement/substitute

19.2 Data Infrastructure

1. **Automated literature monitoring:** NLP to extract new effect estimates
2. **Real-time outcome tracking:** Connect spending to outcomes continuously
3. **API access:** Enable researchers to query OBG data programmatically

19.3 Governance Integration

1. **Dashboard for policymakers:** Real-time gap analysis
2. **Budget proposal scoring:** Automatically assess proposed budgets vs. OSL targets
3. **Incentive Alignment Bonds:** Tie politician compensation to moving toward OSL

20 Conclusion

The Optimal Budget Generator framework provides a systematic, evidence-based approach to budget allocation. Unlike marginal-return frameworks that can justify infinite spending on high-return categories, OBG recognizes that every category has an optimal level - like the Recommended Daily Allowance for nutrients.

The framework answers three questions:

1. **What is the target?** OBG provides evidence-based spending levels for each category
2. **How far are we?** Gap analysis shows where current spending diverges from optimal
3. **How confident are we?** BIS scores evidence quality so policymakers know which OSL estimates are reliable

Even with imperfect evidence, systematically moving from severe misallocation (military 100% above OSL, vaccinations 75% below OSL) toward evidence-based targets will produce welfare gains orders of magnitude larger than current discretionary allocation achieves.

Acknowledgments

The author thanks seminar participants and anonymous reviewers for helpful comments and suggestions. All errors remain the author's own.

21 References

1. Arrow, K. J. *Social Choice and Individual Values*. (Kenneth J. Arrow, 1951).
Arrow's Impossibility Theorem proves that no rank-order voting system can satisfy all of four. "fairness" criteria: unrestricted domain, non-dictatorship, Pareto efficiency, and independence of irrelevant alternatives. This foundational result in social choice theory demonstrates that there is no perfect method for aggregating individual preferences into collective decisions—all voting systems involve tradeoffs. The theorem has profound implications for democratic theory, welfare economics, and mechanism design, showing that preference aggregation is irreducibly political. Additional sources: <https://www.amazon.com/Social-Choice-Individual-Values-Monograph/dp/0300013647>
2. Myerson, R. B. *Optimal auction design*. *Mathematics of Operations Research* **6**, 58–73 (1981)
This seminal paper establishes the Revenue Equivalence Theorem and introduces the "virtual valuation" concept for mechanism design. Myerson shows how to design auctions that maximize expected revenue given incentive-compatible reporting constraints. The paper, along with the Revelation Principle, provides foundational tools for designing mechanisms where agents truthfully report private information. Essential for understanding incentive-compatible oracle design in algorithmic governance systems.
3. Kydland, F. E. & Prescott, E. C. *Rules rather than discretion: The inconsistency of optimal plans*. *Journal of Political Economy* **85**, 473–492 (1977)
Time-inconsistency describes situations where, with the passing of time, policies that were determined to be optimal yesterday are no longer perceived to be optimal today and are not implemented... This insight shifted the focus of policy analysis from the study of individual policy decisions to the design of institutions that mitigate the time consistency problem.
4. CSIS. Smallpox eradication ROI. CSIS <https://www.csis.org/analysis/smallpox-eradication-model-global-cooperation>.
5. Office, C. B. *Estimated impact of the american recovery and reinvestment act on employment and economic output in 2014*. (2015).
6. ICER. ICER QALY methodology and standards. *ICER* <https://icer.org/our-approach/methods-process/cost-effectiveness-the-qaly-and-the-evlyg/> (2024)
The quality-adjusted life year (QALY) is the academic standard for measuring how well all different kinds of medical treatments lengthen and/or improve patients' lives, and therefore the metric has served as a fundamental component of cost-effectiveness analyses in the US and around the world for more than 30 years. ICER's health benefit price benchmark (HBPB) will continue to be reported using the standard range from \$100,000 to \$150,000 per QALY and per evLYG. Additional sources: <https://icer.org/our-approach/methods-process/cost-effectiveness-the-qaly-and-the-evlyg/> | <https://icer.org/wp-content/uploads/2024/02/Reference-Case-4.3.25.pdf>
7. Economic Co-operation, O. for & Development. *OECD government spending as percentage of GDP*. (2024)
OECD government spending data shows significant variation among developed nations: United States: 38.0% of GDP (2023) Switzerland: 35.0% of GDP - 3 percentage points lower than US Singapore: 15.0% of GDP - 23 percentage points lower than US (per IMF data) OECD average: approximately 40% of GDP Additional sources: <https://data.oecd.org/gga/general-government-spending.htm>

8. Papanicolas, I. et al. Health care spending in the united states and other high-income countries. *Papanicolas et al.* <https://jamanetwork.com/journals/jama/article-abstract/2674671> (2018)
The US spent approximately twice as much as other high-income countries on medical care (mean per capita: \$9,892 vs \$5,289), with similar utilization but much higher prices. Administrative costs accounted for 8% of US spending vs 1-3% in other countries. US spending on pharmaceuticals was \$1,443 per capita vs \$749 elsewhere. Despite spending more, US health outcomes are not better. Additional sources: https://jamanetwork.com/journals/jama/article-abstract/2674671
9. PMC. Healthcare investment economic multiplier (1.8). *PMC: California Universal Health Care* <https://pmc.ncbi.nlm.nih.gov/articles/PMC5954824/> (2022)
Healthcare fiscal multiplier: 4.3 (95% CI: 2.5-6.1) during pre-recession period (1995-2007) Overall government spending multiplier: 1.61 (95% CI: 1.37-1.86) Why healthcare has high multipliers: No effect on trade deficits (spending stays domestic); improves productivity & competitiveness; enhances long-run potential output Gender-sensitive fiscal spending (health & care economy) produces substantial positive growth impacts Note: "1.8" appears to be conservative estimate; research shows healthcare multipliers of 4.3 Additional sources: https://pmc.ncbi.nlm.nih.gov/articles/PMC5954824/ | https://cepr.org/voxeu/columns/government-investment-and-fiscal-stimulus | https://ncbi.nlm.nih.gov/pmc/articles/PMC3849102/ | https://set.odg.org/wp-content/uploads/2022/01/Fiscal-multipliers-review.pdf
10. Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. & Yavitz, A. [The rate of return to the HighScope perry preschool program.](#) *Journal of Public Economics* **94**, 114–128 (2010).
11. EPI. Education investment economic multiplier (2.1). *EPI: Public Investments Outside Core Infrastructure* <https://www.epi.org/publication/bp348-public-investments-outside-core-infrastructure/>
Early childhood education: Benefits 12X outlays by 2050; \$8.70 per dollar over life-time Educational facilities: \$1 spent → \$1.50 economic returns Energy efficiency comparison: 2-to-1 benefit-to-cost ratio (McKinsey) Private return to schooling: 9% per additional year (World Bank meta-analysis) Note: 2.1 multiplier aligns with benefit-to-cost ratios for educational infrastructure/energy efficiency. Early childhood education shows much higher returns (12X by 2050) Additional sources: https://www.epi.org/publication/bp348-public-investments-outside-core-infrastructure/ | https://documents1.worldbank.org/curated/en/442521523465644318/pdf/WPS8402.pdf | https://freopp.org/whitepapers/establishing-a-practical-return-on-investment-framework-for-education-and-skills-development-to-expand-economic-opportunity/
12. Batini, N., Di Serio, M., Fragetta, M., Melina, G. & Waldron, A. [Building back better: How big are green spending multipliers?](#) *Ecological Economics* **193**, 107305 (2021).
13. Barro, R. J. & Redlick, C. J. [Macroeconomic effects from government purchases and taxes.](#) *The Quarterly Journal of Economics* **126**, 51–102 (2011).
14. Angrist, J. D. & Pischke, J.-S. [The credibility revolution in empirical economics: How better research design is taking the con out of econometrics.](#) *Angrist & Pischke* **24**, 3–30 (2010)
The primary engine driving improvement has been a focus on the quality of empirical research designs. Additional sources: https://www.aeaweb.org/articles?id=10.1257/jep.24.2.3

15. Jackson, C. K., Johnson, R. C. & Persico, C. [The effects of school spending on educational and economic outcomes: Evidence from school finance reforms](#). *The Quarterly Journal of Economics* **131**, 157–218 (2016)
Using exogenous variation from court-ordered school finance reforms, finds that a 10% increase in per-pupil spending throughout all 12 years of public school leads to 0.31 more completed years of education, 7.25% higher wages, and a 3.67 percentage-point reduction in adult poverty. This is one of the most credible causal estimates of the effect of education spending on adult outcomes, using natural experiments to address reverse causality concerns.
16. Bhutta, D., Z. A. Evidence-based maternal and child health interventions. *Bhutta* [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(13\)60996-4/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(13)60996-4/fulltext) (2014)
Meta-analysis of 34 evidence-based interventions for maternal and child health demonstrates that universal coverage with proven interventions could prevent 2.3 million neonatal deaths and 1.5 million child deaths annually. The benefit-cost ratios for these interventions consistently exceed 10:1.
17. AAF. Return on investment for vaccines. *AAF* <https://www.americanactionforum.org/research/vaccine-protection-and-productivity-the-economic-value-of-vaccines/> (2011)
Every \$1 spent on childhood immunizations results in approximately \$11 in savings. (700% ROI). For low/middle-income countries: \$26.1-\$51.0 ROI using cost-of-illness approach, \$52.2 ROI using value-of-statistical-life approach. US childhood vaccines 1994-2023 saved \$540B in direct costs, \$2.7T in total societal savings. Additional sources: <https://www.americanactionforum.org/research/vaccine-protection-and-productivity-the-economic-value-of-vaccines/> | <https://www.healthaffairs.org/doi/10.1377/hlthaff.2020.00103> | https://immunizationevidence.org/featured_issues/the-value-of-vaccines-investments-in-immunization-yield-high-returns/
18. CDC. Childhood vaccination (US) ROI. *CDC* <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6316a4.htm> (2017).
19. MMWR, C. Childhood vaccination economic benefits. *CDC MMWR* <https://www.cdc.gov/mmwr/volumes/73/wr/mm7331a2.htm> (1994)
US programs (1994-2023): \$540B direct savings, \$2.7T societal savings (\$18B/year direct, \$90B/year societal) Global (2001-2020): \$820B value for 10 diseases in 73 countries (\$41B/year) ROI: \$11 return per \$1 invested Measles vaccination alone saved 93.7M lives (61% of 154M total) over 50 years (1974-2024) Additional sources: <https://www.cdc.gov/mmwr/volumes/73/wr/mm7331a2.htm> | [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(24](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)
20. Bloom, N., Schankerman, M. & Van Reenen, J. [Identifying technology spillovers and product market rivalry](#). *Econometrica* **81**, 1347–1393 (2013)
Estimates social returns to R&D of 20-60% through technology spillovers, substantially exceeding private returns. Provides empirical evidence for the positive externalities of research investment, supporting public funding of basic research.
21. GiveWell. Cost per DALY for deworming programs. <https://www.givewell.org/international/technical/programs/deworming/cost-effectiveness>
Schistosomiasis treatment: \$28.19-\$70.48 per DALY (using arithmetic means with varying disability weights) Soil-transmitted helminths (STH) treatment: \$82.54 per DALY (mid-point estimate) Note: GiveWell explicitly states this 2011 analysis is "out of date" and their current methodology focuses on long-term income effects rather than short-term health DALYs Additional sources: <https://www.givewell.org/international/technical/programs/deworming/cost-effectiveness>

22. Abadie, A., Diamond, A. & Hainmueller, J. [Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program](#). *Journal of the American Statistical Association* **105**, 493–505 (2010)
The synthetic control method provides a systematic way to choose comparison units in comparative case studies. A combination of comparison units often provides a better comparison for the unit affected by the policy intervention than any single comparison unit alone.
23. Institute, S. I. P. R. [Trends in world military expenditure, 2023](#). (2024).
24. Institute, S. I. P. R. [Trends in world military expenditure, 2024](#). (2024)
NATO members spent \$1.506 trillion in 2024 (55% of world military spending).. European NATO spent \$454 billion. US spent \$968 billion. Additional sources: https://www.sipri.org/sites/default/files/2025-04/2504_fs_milex_2024.pdf
25. Mercatus. Military spending economic multiplier (0.6). *Mercatus: Defense Spending and Economy* <https://www.mercatus.org/research/research-papers/defense-spending-and-economy>
Ramey (2011): 0.6 short-run multiplier Barro (1981): 0.6 multiplier for WWII spending (war spending crowded out 40¢ private economic activity per federal dollar) Barro & Redlick (2011): 0.4 within current year, 0.6 over two years; increased govt spending reduces private-sector GDP portions General finding: \$1 increase in deficit-financed federal military spending = less than \$1 increase in GDP Variation by context: Central/Eastern European NATO: 0.6 on impact, 1.5-1.6 in years 2-3, gradual fall to zero Ramey & Zubairy (2018): Cumulative 1% GDP increase in military expenditure raises GDP by 0.7% Additional sources: <https://www.mercatus.org/research/research-papers/defense-spending-and-economy> | <https://cepr.org/voxeu/columns/world-war-ii-america-spending-deficits-multipliers-and-sacrifice> | https://www.rand.org/content/dam/rand/pubs/research_reports/RRA700/RRA739-2/RAND_RRA739-2.pdf

22 Appendix A: Worked Example - Complete OBG Calculation

22.1 Example: US Military Discretionary Spending

This worked example demonstrates the complete OBG calculation for a category where reference benchmarking is the primary method. Military spending data comes from the Stockholm International Peace Research Institute (SIPRI), which maintains the most comprehensive global military expenditure database²³.

Step 1: Define the category

- **Category:** Military discretionary spending (defense budget excluding veterans’ benefits and military pensions)
- **Current US spending:** \$850B (FY2024)
- **Outcome of interest:** National security (deterrence, territorial integrity)

Step 2: Select reference countries

Reference data from SIPRI Military Expenditure Database²⁴:

| Country | Military % GDP (2023) | GDP (trillion USD) | Selection criteria |
|-----------|-----------------------|--------------------|--|
| Germany | 1.5% | \$4.1T | NATO member, high-income, strong institutions |
| France | 1.9% | \$2.8T | NATO member, nuclear power, high-income |
| UK | 2.2% | \$3.1T | NATO member, nuclear power, high-income |
| Japan | 1.0% | \$4.2T | High-income, strong institutions, regional threats |
| Australia | 2.1% | \$1.7T | High-income, alliance partner |
| Canada | 1.3% | \$2.1T | NATO member, neighbor |

Median reference: 1.7% of GDP (median of: 1.0%, 1.3%, 1.5%, 1.9%, 2.1%, 2.2%)

Step 3: Context adjustment

| Factor | Adjustment | Rationale |
|-----------------------|------------------|--|
| Global role | +0.5% | US provides NATO umbrella |
| Geographic security | -0.3% | US has oceanic borders, friendly neighbors |
| Existing alliances | -0.2% | Cost-sharing with allies |
| Nuclear deterrent | Already included | Reference countries include nuclear powers |
| Net adjustment | +0.0% | Adjustments roughly cancel |

Step 4: Calculate OSL

- US GDP: \$27T
- Reference spending: 1.7% of GDP
- Adjustment: 0%
- **OSL = 1.7% × \$27T = \$459B**

Step 5: Gap analysis

| Metric | Value |
|------------------|---------------------------------|
| Current spending | \$850B |
| OSL | \$459B |
| Gap | -\$391B (overinvestment) |
| Gap % of current | -46% |

Step 6: Evidence assessment

| Criterion | Assessment | Score |
|-------------------------------|------------------------------------|-------------|
| Reference country consistency | Moderate (1.0-2.2% range) | 0.6 |
| Context transferability | Uncertain (US global role unique) | 0.4 |
| Outcome linkage | Weak (spending → security unclear) | 0.3 |
| Alternative methods | Limited | 0.4 |
| BIS | | 0.50 |

Evidence grade: C (Mixed evidence - benchmark clear, but US context unique)

Step 7: Multi-unit reporting

| Unit Level | Value | Interpretation |
|-------------------------|-----------------|--|
| Natural | -\$391B gap | 46% overinvestment vs. peers |
| Per capita | -\$1,170/person | Americans pay \$2,550 vs. \$1,380 peer avg |
| Opportunity cost | 4-10x | Returns if reallocated to high-return categories |
| Composite (BIS) | 0.50 | Moderate confidence in OSL estimate |

Recommendation: Strong evidence of overinvestment relative to peer countries. The fiscal multiplier for military spending is estimated at 0.6-0.8, lower than most domestic programs²⁵. However, US global role creates genuine uncertainty about context transferability. Recommend gradual reduction (10% per year) with continuous outcome monitoring.

23 Appendix B: Analysis Workflow

23.1 Complete OBG Analysis Pipeline

```

+-----+
|               OBG ANALYSIS WORKFLOW               |
+-----+
```

Phase 1: DATA COLLECTION

1. Budget data ingestion
 - Pull current spending by category (OMB, USASpending)
 - Normalize categories to standard taxonomy
 - Identify subcategories for detailed analysis
 - Flag data quality issues
2. Reference country data
 - Pull spending data from OECD, World Bank
 - Filter by reference country criteria
 - Normalize to per-capita and % GDP
 - Calculate medians and distributions
3. Effect estimate data

- +-- Search systematic reviews and meta-analyses
- +-- Extract effect sizes with standard errors
- +-- Code study quality (RCT, natural experiment, etc.)
- +-- Build literature database by category

Phase 2: OBG ESTIMATION

- 4. Reference benchmarking
 - +-- Calculate median reference spending
 - +-- Apply context adjustments
 - +-- Estimate OSL with confidence intervals
 - +-- Document methodology
- 5. Diminishing returns modeling (where data permits)
 - +-- Fit nonlinear spending-outcome functions
 - +-- Identify "knee" of curve
 - +-- Calculate marginal returns at current spending
 - +-- Estimate optimal level
- 6. Cost-effectiveness analysis (health/life-saving)
 - +-- Identify interventions below CE threshold
 - +-- Calculate scale-up costs
 - +-- Sum to category OSL
 - +-- Document assumptions
- 7. Method reconciliation
 - +-- Compare OSL estimates across methods
 - +-- Weight by method reliability
 - +-- Produce consensus OSL estimate
 - +-- Flag discrepancies

Phase 3: EVIDENCE QUALITY

- 8. BIS calculation
 - +-- Compute quality weights per study
 - +-- Compute precision weights
 - +-- Compute recency weights
 - +-- Aggregate to category BIS
- 9. Evidence grading
 - +-- Assign A-F grade based on BIS
 - +-- Document key evidence
 - +-- Identify research gaps
 - +-- Flag high-uncertainty categories

Phase 4: GAP ANALYSIS

- 10. Compute gaps

```

+--- Gap = OSL - Current
+--- Calculate % gap
+--- Classify as under/over/optimal
+--- Apply BIS weighting

11. Priority ranking
+--- Priority = |Gap| × BIS
+--- Rank categories
+--- Identify reallocation pairs
+--- Estimate welfare gains

Phase 5: OUTPUT GENERATION
-----

12. Multi-unit reporting
+--- Natural units ($/capita, % GDP)
+--- Monetized (ROI, opportunity cost)
+--- Health units (QALYs where applicable)
+--- Composite (BIS, evidence grade)

13. Sensitivity analysis
+--- Vary key parameters
+--- Test reference country sets
+--- Report OSL ranges
+--- Identify robust conclusions

14. Documentation
+--- Generate category reports
+--- Create methodology audit trail
+--- Version control estimates
+--- Publish to dashboard/API

```

24 Appendix C: Glossary

24.1 Core Concepts

- **Optimal Budget Generator (OBG):** The framework/methodology for generating integrated budget recommendations based on evidence of spending-outcome relationships. OBG accounts for the zero-sum nature of budget allocation and produces Optimal Spending Level (OSL) estimates for each category.
- **Optimal Spending Level (OSL):** The evidence-based target spending level for each category, produced by the OBG framework. OSL_i represents the optimal spending level for category i . Below OSL indicates underinvestment; above OSL indicates diminishing returns.
- **Budget Impact Score (BIS):** A 0-1 score measuring confidence in each category's OSL estimate based on the quality and quantity of causal evidence. Higher BIS indicates more reliable OSL recommendations.
- **Spending Gap:** The difference between current spending and the evidence-based target for

each category. Positive gaps indicate underinvestment; negative gaps indicate overinvestment.

- **Reference Country Benchmarking:** Estimating target spending levels by observing spending in comparable high-performing countries and adjusting for context.
- **Diminishing Returns:** The economic principle that marginal returns to spending decrease as spending increases. The optimal level is where marginal return equals opportunity cost.

24.2 Estimation Methods

- **Context Adjustment:** Modifications to reference country benchmarks accounting for differences in population, geography, institutions, and existing infrastructure.
- **Cost-Effectiveness Threshold:** The maximum acceptable cost per QALY (or other health outcome) for including an intervention in target calculations. Typically \$50K-\$150K per QALY.
- **Dose-Response Curve:** The relationship between spending level (dose) and outcome (response). Used to identify diminishing returns and estimate optimal spending levels.

24.3 Evidence Quality

- **Quality Weight (w^Q):** Weight assigned to a study based on identification strategy. RCTs receive 1.0; cross-sectional studies receive 0.25.
- **Precision Weight (w^P):** Weight assigned based on standard error. More precise estimates receive higher weight.
- **Recency Weight (w^R):** Weight assigned based on publication date. More recent studies receive higher weight via exponential decay.
- **Evidence Grade:** Letter grade (A-F) summarizing confidence in each category's target estimate. A = strong evidence; F = insufficient evidence.

24.4 Output Concepts

- **Priority Score:** Product of gap magnitude and BIS. Used to rank categories for reallocation priority.
- **Value of Information (VOI):** Expected benefit of additional research on uncertain categories. High-VOI categories warrant pilot funding.
- **Multi-Unit Reporting:** Presenting results in natural units, monetized equivalents, health units, and composite scores for interpretability.

25 Appendix D: Comparison to Actual US Budget

25.1 Current US Discretionary Budget vs. OSL Targets

| Category | Current (B) OSL(B) | Gap (\$B) | Gap % | BIS | Prior- ity | |
|-------------------------------|-----------------------|-----------|-------|-------|---------------|-----|
| Defense (discretionary) | 850 | 459 | -391 | -46% | 0.50 | 195 |
| Non-defense discretionary | 915 | 1,300 | +385 | +42% | 0.65 | 250 |
| - <i>Education</i> | 80 | 120 | +40 | +50% | 0.75 | 30 |
| - <i>Health</i> (research) | 50 | 100 | +50 | +100% | 0.80 | 40 |
| - <i>Vaccinations</i> | 8 | 35 | +27 | +338% | 0.95 | 26 |
| - <i>Basic</i> research | 45 | 90 | +45 | +100% | 0.70 | 32 |
| - <i>Infrastructure</i> | 100 | 150 | +50 | +50% | 0.60 | 30 |
| - <i>Early</i> childhood | 50 | 70 | +20 | +40% | 0.85 | 17 |
| Agricultural subsidies | 25 | 0 | -25 | -100% | 0.90 | 23 |

Key findings:

1. **Severe overinvestment:** Military spending is ~85% above reference benchmarks
2. **Severe underinvestment:** Vaccinations, basic research, health research far below evidence-optimal levels
3. **Negative-return spending:** Agricultural subsidies should be eliminated entirely
4. **Reallocation potential:** ~\$400B could be reallocated from low/negative return to high-return categories

Estimated welfare gain from OSL alignment: Moving from current allocation to OSL targets would increase welfare-equivalent output by an estimated 3-5% of GDP (\$750B-\$1.25T annually), based on the differential returns between over- and under-invested categories.

Corresponding Author: Mike P. Sinn, Decentralized Institutes of Health (mike@warondisease.org)

Conflicts of Interest: The author declares no conflicts of interest.

Funding: This work received no external funding.

Data Availability: All data sources referenced in this paper are publicly available: OECD iLibrary (education, health spending), World Bank WDI (cross-country indicators), SIPRI Military Expenditure Database (defense spending), and CDC vaccination cost data. URLs are provided in the Data Sources section. A complete replication package including analysis code, data extraction scripts, and worked example calculations will be deposited in a public repository (GitHub/Zenodo) upon publication.

Ethics Statement: This is a methodological specification. No human subjects research was conducted.

Preprint: This working paper has not undergone peer review.