The Continuous Evidence Generation Protocol: Two-Stage Validation (RWE → Pragmatic Trials)

Mike P. Sinn

Institute for Accelerated Medicine

m@warondisease.org | ORCID: 0009-0006-0212-1094

Online version: https://manual.warondisease.org/knowledge/appendix/dfda-spec-paper.html

Abstract

Treatments that could save lives take an average of 8.2 years (95% CI: 4.85 years-11.5 years) to complete clinical trials after discovery. Since 1962, these delays have contributed to an estimated 102 million deaths (95% CI: 36.9 million deaths-214 million deaths) preventable deaths. Meanwhile, only 1-10% of adverse drug events get reported to the FDA, and billions of people generate continuous health data through wearables and apps that remains unharvested.

We present a two-stage framework that transforms this data into validated treatment recommendations. Stage 1 ($0.1 (95% CI: $0.03-$1)/patient): aggregate millions of natural experiments and score causal confidence using the Predictor Impact Score (PIS), a composite metric operationalizing six Bradford Hill causality criteria. Stage 2 ($929 (95% CI: $97-$3K)/patient): confirm top signals through pragmatic trials embedded in routine care, 44.1x (95% CI: 39.4x-89.1x) cheaper than traditional Phase III trials. Cost estimates derive from a meta-analysis of 108 pragmatic trials plus implementations like RECOVERY (which found a life-saving treatment in 100 days) and ADAPTABLE. A Trial Priority Score (PIS x DALYs x Novelty x Feasibility) determines which signals proceed to experimental confirmation.

The framework produces three outputs absent from current pharmacovigilance: (1) "Outcome Labels," per-condition documents ranking all treatments by quantitative effect size (inverting the traditional per-drug FDA label paradigm); (2) precision dosing recommendations derived from optimal daily values (the predictor values historically preceding the best outcomes); and (3) a three-tier evidence grading system (Validated, Promising, Signal) combining observational and experimental effect sizes. Trial results feed back to calibrate observational models, creating a learning health system where accuracy improves continuously.

High PIS signals warrant experimental investigation; low PIS does not rule out true effects. This framework complements traditional RCTs. Stage 2 pragmatic trials are required to establish validated causal claims.

Table of contents

1 Abstract **5**

2 System Overview: From Methodology to Implementation **7**

2.1 What Patients See 8

2.2 What Companies See 10

2.3 Where This Methodology Fits 11

3 Introduction **12**
3.1 The Human Cost of the Current System . . . 12
3.2 The Pharmacovigilance Gap . . . 14
3.3 The Real-World Data Opportunity . . . 16
3.4 Our Contribution . . . 17

4 Data Collection and Integration **17**
4.1 Data Sources . . . 17
4.2 Variable Ontology . . . 18
4.3 Measurement Structure . . . 21
4.4 Unit Standardization . . . 22

5 Mathematical Framework **22**
5.1 Data Structure . . . 23
5.2 Temporal Alignment . . . 24
5.2.1 Onset Delay and Duration of Action . . . 24
5.2.2 Outcome Window Calculation . . . 25
5.3 Pair Generation Strategies . . . 25
5.3.1 Outcome-Based Pairing (Predictor has Filling Value) . . . 26
5.3.2 Predictor-Based Pairing (No Filling Value) . . . 28
5.4 Filling Value Logic . . . 29
5.4.1 Filling Types . . . 29
5.4.2 Temporal Boundaries . . . 29
5.4.3 Conservative Bias . . . 31
5.5 Baseline Definition and Outcome Estimation . . . 32
5.5.1 Within-Subject Comparison . . . 32
5.5.2 Outcome Means . . . 32
5.6 Percent Change from Baseline . . . 32
5.7 Correlation Coefficients . . . 32
5.7.1 Pearson Correlation (Linear Relationships) . . . 32
5.7.2 Spearman Rank Correlation (Monotonic Relationships) . . . 33
5.7.3 Forward and Reverse Correlations . . . 33
5.8 Z-Score Normalization . . . 34
5.9 Statistical Significance . . . 34
5.10 Hyperparameter Optimization . . . 34

6 Population Aggregation **35**
6.1 Individual to Population . . . 35
6.2 Standard Error and Confidence Intervals . . . 35
6.3 Heterogeneity Assessment . . . 35

7 Data Quality Requirements **35**
7.1 Minimum Thresholds . . . 35
7.2 Variance Validation . . . 36
7.3 Outcome Value Spread . . . 36

8 Predictor Impact Score **36**
8.1 What Makes the Predictor Impact Score Novel . . . 38

8.2 User-Level Predictor Impact Score . . . 38
8.3 Aggregate (Population-Level) Predictor Impact Score . . . 38
8.4 Z-Score and Effect Magnitude Factor . . . 39
8.5 Temporality Factor . . . 39
8.6 Percent Change from Baseline . . . 39
8.7 Statistical Significance . . . 40
8.8 Interest Factor . . . 40
8.9 Additional Data Quality Components . . . 40
8.10 Bradford Hill Criteria Mapping . . . 41
8.11 Interpreting Predictor Impact Scores . . . 41

9 Provisional Thresholds - Not Yet Validated 41
9.1 Optimal Daily Value for Precision Dosing . . . 42
9.1.1 Value Predicting High Outcome . . . 43
9.1.2 Value Predicting Low Outcome . . . 43
9.1.3 Grouped Optimal Values . . . 43
9.1.4 Precision Dosing Recommendations . . . 44
9.1.5 Mathematical Relationship to Biological Gradient . . . 45
9.1.6 Clinical Applications . . . 46
9.1.7 Limitations . . . 46
9.1.8 Confidence Intervals for Optimal Values . . . 47
9.1.9 Individual vs Population Optimal Values . . . 48
9.1.10 Temporal Stability and Recalculation . . . 48
9.1.11 Edge Cases: Minimal Dose-Response . . . 49
9.1.12 Validation of Optimal Values . . . 49
9.2 Saturation Constant Rationale . . . 50
9.3 Effect Following High vs Low Predictor Values . . . 50
9.3.1 Average Outcome Metrics . . . 50
9.3.2 Calculation . . . 51
9.4 Predictor Baseline and Treatment Averages . . . 51
9.5 Relationship Quality Filters . . . 52
9.5.1 Filter Flags . . . 52
9.5.2 Boring Relationship Definition . . . 53
9.5.3 Usefulness and Causality Voting . . . 54
9.6 Variable Valence . . . 54
9.6.1 Impact on Interpretation . . . 54
9.7 Temporal Parameter Optimization . . . 54
9.7.1 Stored Optimization Data . . . 54
9.7.2 Optimization Grid . . . 55
9.7.3 Overfitting Protection . . . 57
9.8 Spearman Rank Correlation . . . 58

10 Outcome Label Generation 59
10.1 Predictor Analysis Reports . . . 59
10.2 Report Structure . . . 60
10.3 Category-Specific Analysis . . . 61
10.4 Verification Status . . . 62
10.5 Outcome Labels vs. FDA Drug Labels . . . 62

10.6 Worked Example: Complete Outcome Label . . . 62

11 Treatment Ranking System **65**
11.1 Within-Category Rankings . . . 65
11.2 Ranking Algorithm . . . 66
11.3 Confidence Weighting . . . 66
11.4 Comparative Effectiveness Display . . . 66

12 Safety and Efficacy Quantification **66**
12.1 Safety Signal Detection . . . 66
12.2 Efficacy Signal Detection . . . 67
12.3 Benefit-Risk Assessment . . . 67

13 Addressing the Bradford Hill Criteria **67**
13.1 Complete Criteria Mapping . . . 68
13.2 Quantitative Criteria Details . . . 69

14 Validation and Quality Assurance **69**
14.1 User Voting System . . . 69
14.2 Automated Quality Checks . . . 70
14.3 Flagged Study Handling . . . 71

15 Stage 2: Pragmatic Trial Confirmation **72**
15.1 The Two-Stage Pipeline . . . 73
15.2 Pragmatic Trial Methodology . . . 74
15.3 Signal-to-Trial Prioritization . . . 75
15.4 Comparative Effectiveness Randomization . . . 75
15.5 Feedback Loop: Trial Results Improve Observational Models . . . 75
15.6 Output: Validated Outcome Labels . . . 77

16 Limitations and How They're Addressed **77**
16.1 Fundamental Limitations: Observational Stage . . . 77
16.2 Methodological Weaknesses: Addressed by Two-Stage Design . . . 78
16.3 Residual Limitations . . . 78
16.4 What This Framework CAN Now Do . . . 79

17 Implementation Guide **80**
17.1 System Architecture . . . 80
17.2 Core Algorithm: Pair Generation . . . 81
17.3 Core Algorithm: Baseline Separation . . . 82
17.4 Algorithm 3: Predictor Impact Score Calculation . . . 82
17.5 Reference Implementation . . . 83

18 Regulatory Considerations **84**
18.1 Positioning Relative to RCTs . . . 84
18.2 Evidence Hierarchy Integration . . . 85
18.3 FDA Real-World Evidence Framework Alignment . . . 86

19 Validation Framework **87**

19.1 The Critical Question . . . 87
19.2 Proposed Validation Study . . . 88
19.3 Known Limitations Requiring Validation . . . 89

20 Future Directions **90**
20.1 Methodological Improvements . . . 90
20.2 Validation Priorities . . . 92
20.3 Implementation Enhancements . . . 93

21 Conclusion **93**

22 Appendix A: Effect Size Classification **94**

23 Appendix B: Variable Category Defaults **94**

24 Appendix C: Glossary **94**

25 Appendix D: Worked Example **96**
25.1 Example: Calculating Predictor Impact Score for "Magnesium → Sleep Quality" . . . 96

26 Appendix E: Analysis Workflow **98**

Version note. This document is an archival snapshot. Latest interactive version: https://manual.warondisease.org/knowledge/appendix/dfda-spec-paper.html. Citable DOI: https://doi.org/10.5281/zenodo.18203375.

1 Abstract

Current pharmacovigilance systems rely primarily on spontaneous adverse event reporting, which suffers from significant underreporting, lack of denominator data, and inability to quantify effect sizes. Meanwhile, the proliferation of wearable devices, health apps, and patient-reported outcomes has generated unprecedented volumes of longitudinal real-world data (RWD) that remain largely untapped for safety and efficacy signal detection.

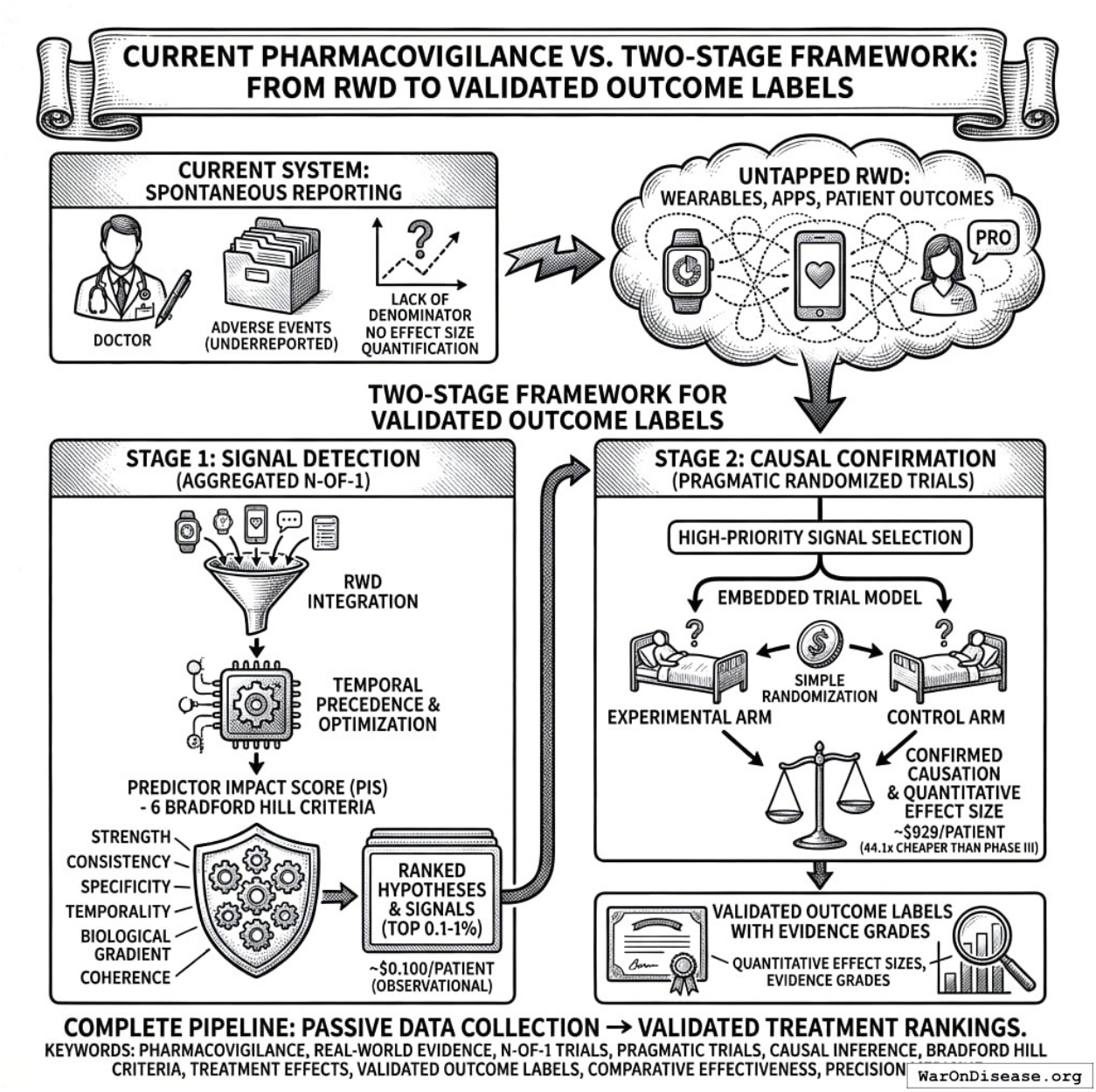

Figure 1: Step 1: Let computers watch a billion people take medicine. Step 2: Test the interesting bits. You were doing Step 2 first, which is why everything costs a billion dollars.

We present a comprehensive **two-stage framework** for generating **validated outcome labels** with quantitative effect sizes:

Stage 1 (Signal Detection): Aggregated N-of-1 observational analysis[139,140] integrates data from millions of individual longitudinal natural experiments. The methodology applies temporal precedence analysis with automated hyperparameter optimization, addresses six of nine Bradford Hill causality criteria through a composite Predictor Impact Score (PIS), and produces ranked treatment-outcome hypotheses at ~$0.1 (95% CI: $0.03-$1).

Stage 2 (Causal Confirmation): High-priority signals (top 0.1-1% by PIS) proceed to pragmatic

randomized trials following the embedded trial model validated across 108+ studies[1,141]. Simple randomization embedded in routine care confirms causation at ~$929 (95% CI: $97-$3K) (44.1x (95% CI: 39.4x-89.1x) cheaper than traditional Phase III trials) while eliminating confounding concerns inherent in observational data.

The complete methodology includes: (1) data collection from heterogeneous sources; (2) temporal alignment with onset delay optimization; (3) within-subject baseline/follow-up comparison; (4) Predictor Impact Score calculation operationalizing Bradford Hill criteria; (5) Trial Priority Score for signal-to-trial prioritization; (6) pragmatic trial protocols for causal confirmation; and (7) validated outcome label generation with evidence grades.

This two-stage design addresses the fundamental limitations of purely observational pharmacovigilance (confounding, self-selection, and inability to prove causation) while maintaining the scale and cost advantages of real-world data. The result is a complete pipeline from passive data collection to validated treatment rankings, presented as both scientific methodology and implementation blueprint for next-generation regulatory systems.

Keywords: pharmacovigilance, real-world evidence, N-of-1 trials, pragmatic trials, causal inference, Bradford Hill criteria, treatment effects, validated outcome labels, comparative effectiveness, precision medicine

2 System Overview: From Methodology to Implementation

This paper describes the statistical methodology powering a patient-facing system best understood as **"Consumer Reports for drugs"** - a searchable database where patients can look up any condition and see every treatment ranked by real-world effectiveness, with quantitative outcome labels showing exactly what happened to people who tried each option.

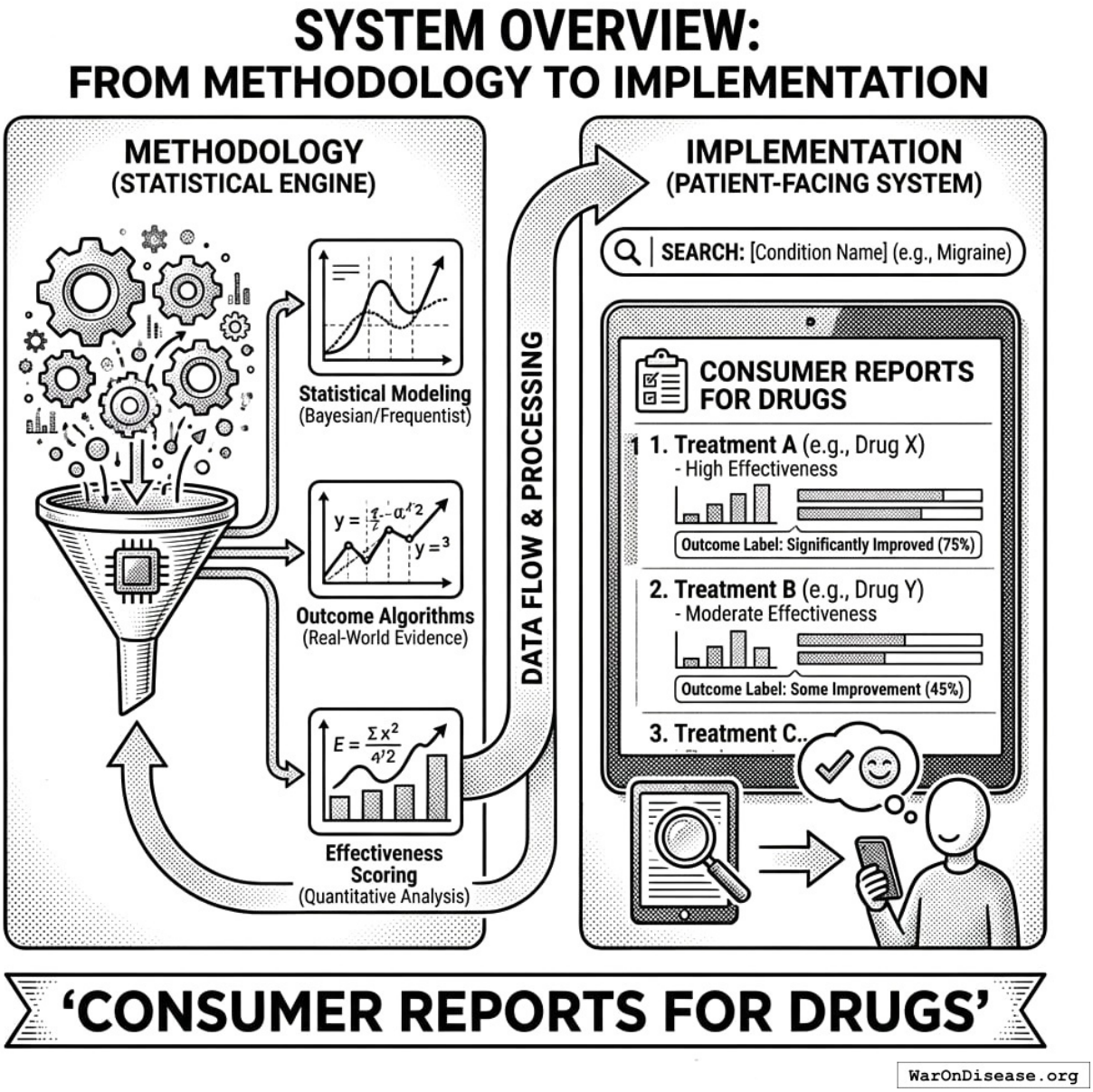

Figure 2: Imagine if restaurants had to tell you which dishes actually taste good instead of just not poisoning you. This is that, but for drugs.

2.1 What Patients See

When a patient searches for their condition, they see:

1. **Treatment Rankings**: Every option (FDA-approved drugs, supplements, lifestyle interventions, experimental treatments) ranked by effect size from real patient data
2. **Outcome Labels**: "Nutrition facts for drugs" showing percent improvement, side effect rates, and sample sizes - not marketing claims
3. **Trial Access**: One-click enrollment in available trials, from home, via any device
4. **Personalized Predictions**: Based on their health data, which treatments work best for

people like them

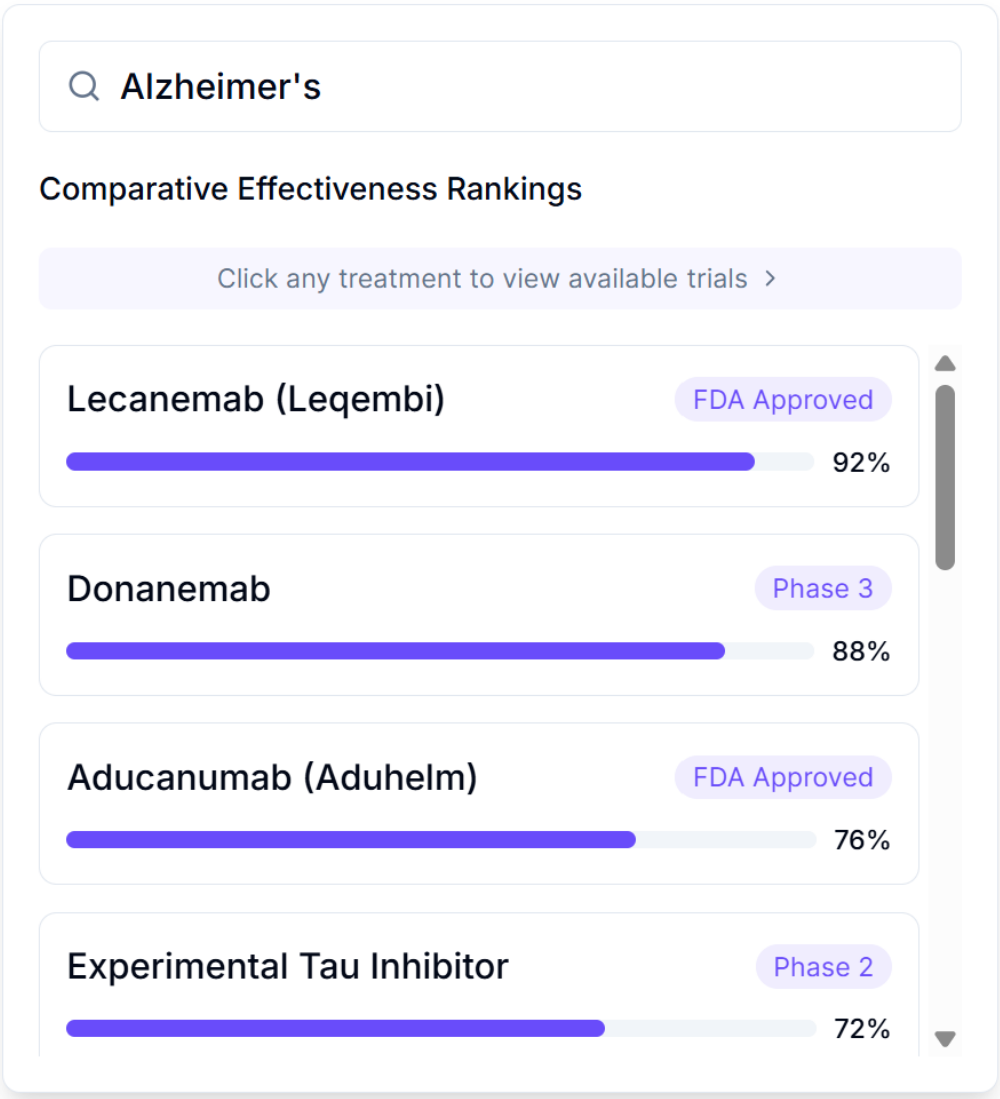

Figure 3: Treatment rankings, like Yelp reviews, but for not dying. You could have done this decades ago. You chose not to.

2.2 What Companies See

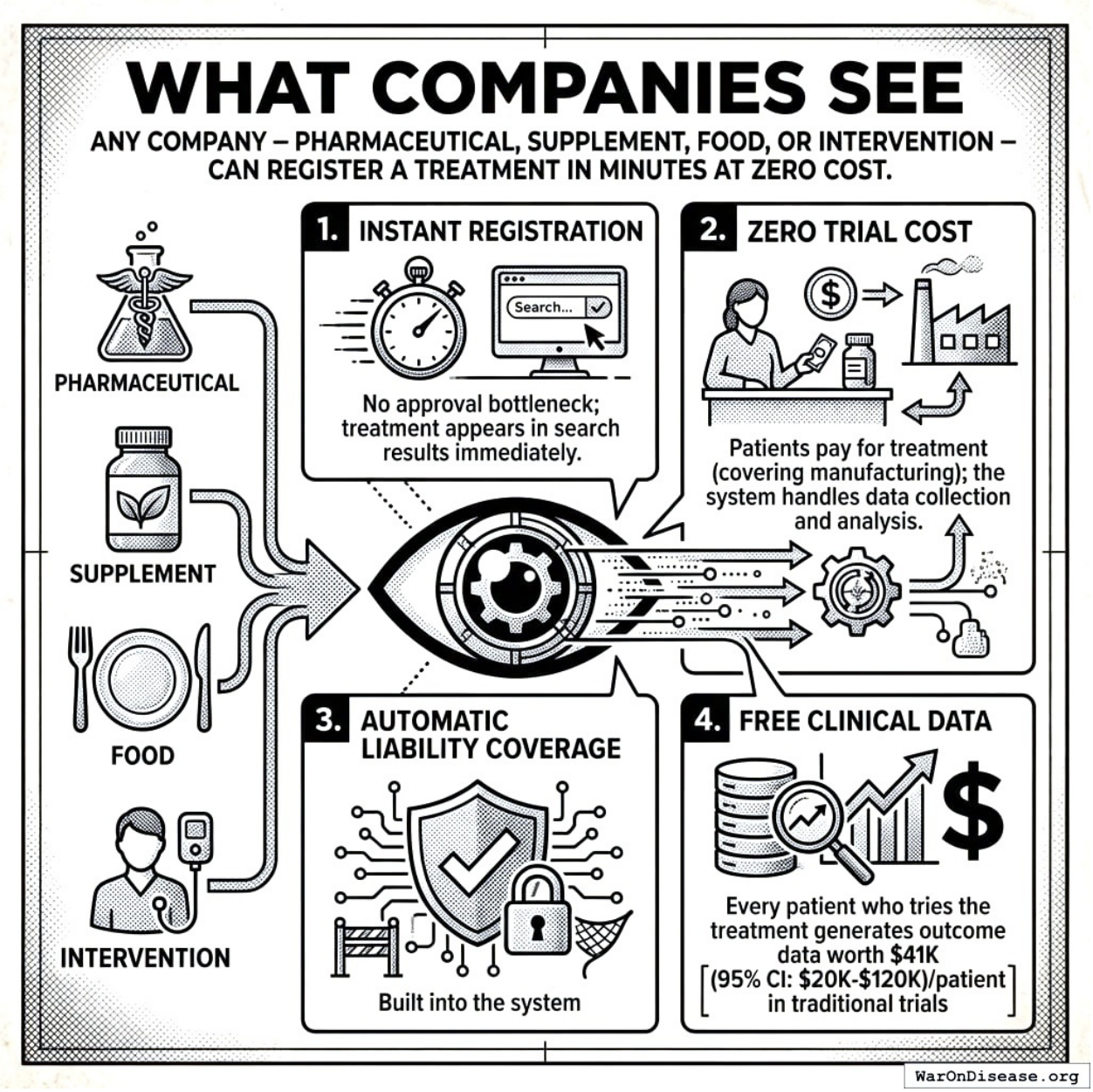

Figure 4: Drug companies used to spend ten years asking permission to help people. Now they just help people and write down what happens. Revolutionary.

Any company - pharmaceutical, supplement, food, or intervention - can register a treatment in minutes at zero cost:

1. **Instant Registration**: No approval bottleneck; treatment appears in search results immediately
2. **Zero Trial Cost**: Patients pay for treatment (covering manufacturing); the system handles data collection and analysis
3. **Automatic Liability Coverage**: Built into the system

4. **Free Clinical Data**: Every patient who tries the treatment generates outcome data worth $41K (95% CI: $20K-$120K)/patient in traditional trials

2.3 Where This Methodology Fits

The **Predictor Impact Score (PIS)** described in this paper is the engine that powers treatment rankings. It transforms raw patient data into the ranked, quantified outcome labels that patients and clinicians use to make decisions. The two-stage pipeline ensures that:

- **Stage 1** (this methodology) generates treatment rankings from millions of real-world observations at ~$0.1 (95% CI: $0.03-$1)/patient
- **Stage 2** (pragmatic trials) confirms causation for high-priority signals at ~$929 (95% CI: $97-$3K)/patient

The result is a self-improving system where every patient's experience helps the next patient make better decisions, transforming the current bottleneck of 1.9 million patients/year (95% CI: 1.5 million patients/year-2.3 million patients/year) annual trial participants into a system where anyone can contribute to medical knowledge.

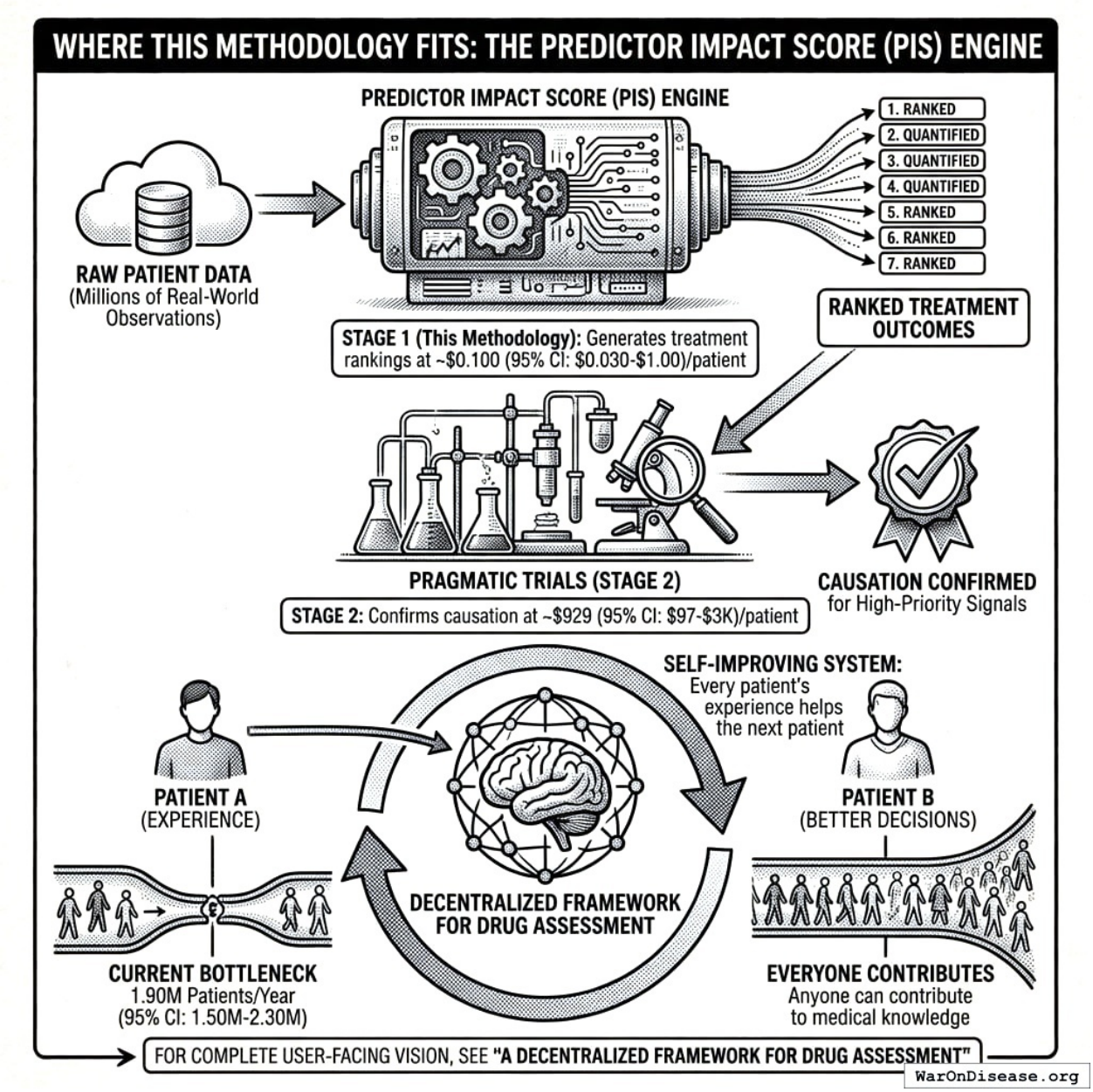

Figure 5: First, computers find patterns in real life. Then, humans check if the computers are hallucinating. It's like peer review, but one of the peers is a billion people.

For the complete user-facing vision, see A Decentralized Framework for Drug Assessment.

3 Introduction

3.1 The Human Cost of the Current System

Every year, 55 million (95% CI: 46.6 million-63.2 million) people die from diseases for which treatments exist or could exist. The tragedy is not that we lack medical knowledge. It's that our system for generating and validating that knowledge operates at a fraction of its potential capacity.

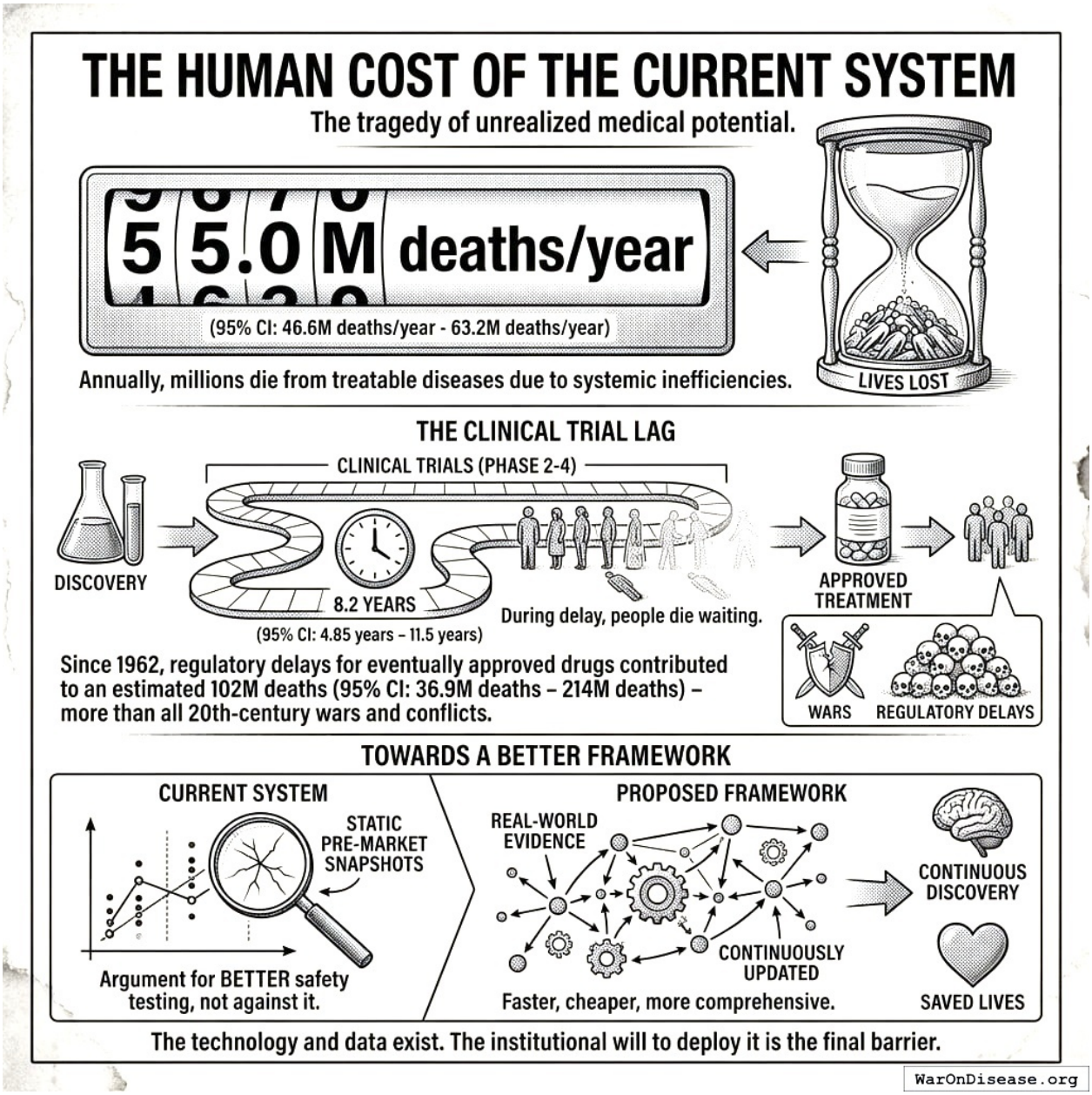

Figure 6: While you waited for permission to try new cancer drugs, more people died than in all of World War II. The forms were very thorough though.

Consider: a treatment that could save lives today takes an average of 8.2 years (95% CI: 4.85 years-11.5 years) to complete Phase 2-4 clinical trials after initial discovery. During this delay, people die waiting. Since 1962, regulatory testing delays for drugs that were eventually approved have contributed to an estimated 102 million deaths (95% CI: 36.9 million deaths-214 million deaths) preventable deaths, more than all wars and conflicts of the 20th century combined.

This is not an argument against safety testing. It is an argument for *better* safety testing: faster, cheaper, more comprehensive, and continuously updated with real-world evidence rather than static pre-market snapshots.

The framework presented here could eliminate this efficacy lag for existing treatments while simultaneously enabling continuous discovery of new therapeutic relationships. The technology exists. The data exists. What remains is the institutional will to deploy it.

3.2 The Pharmacovigilance Gap

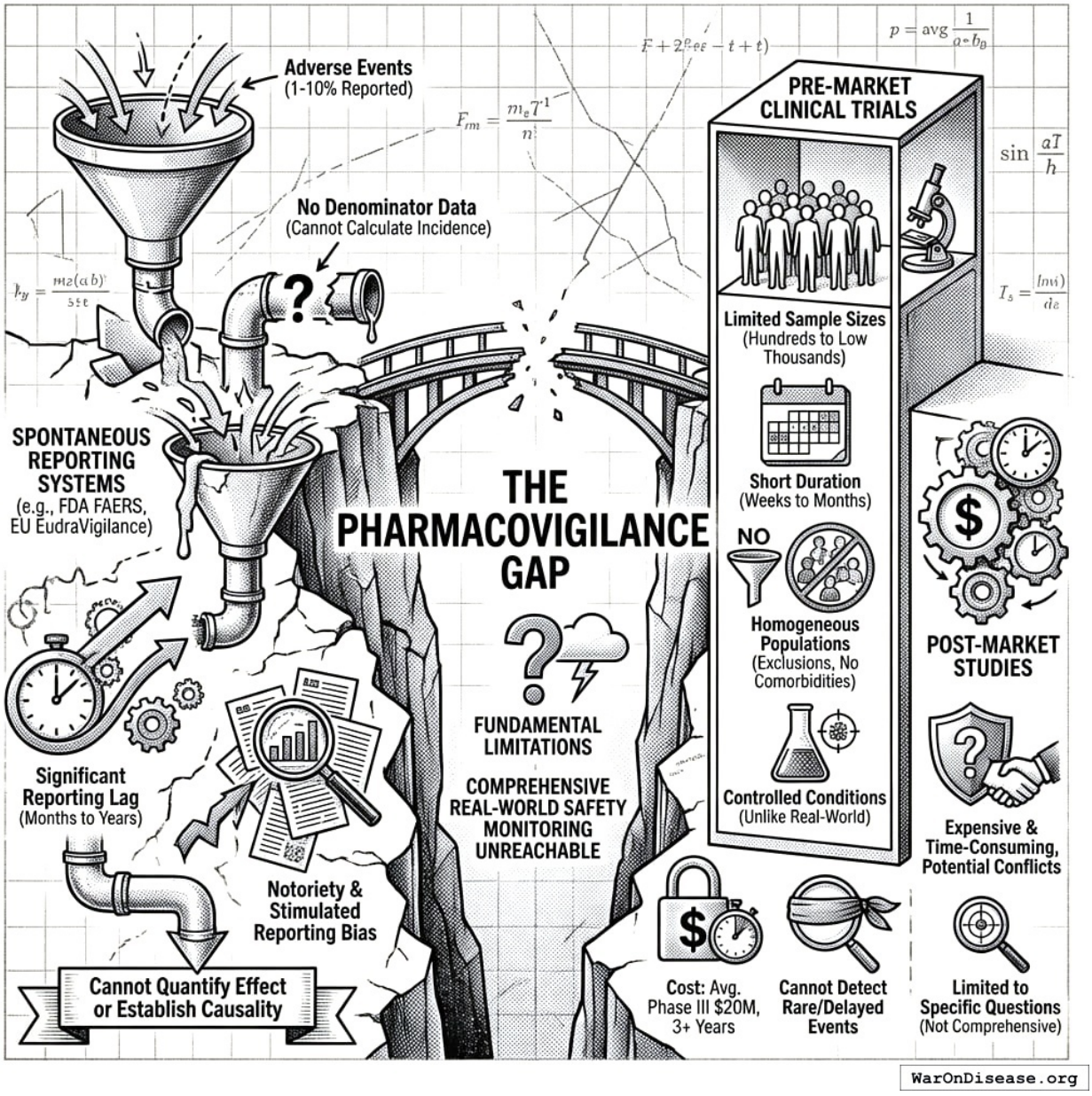

Figure 7: Your three ways of checking if drugs kill people: slow and expensive, slower and more expensive, or fast but everyone lies on the survey.

Modern pharmacovigilance (the science of detecting, assessing, and preventing adverse effects of pharmaceutical products) faces fundamental limitations:

Spontaneous Reporting Systems (e.g., FDA FAERS, EU EudraVigilance):

- Estimated 1-10% of adverse events are reported[142]
- No denominator data (cannot calculate incidence rates)
- Cannot quantify effect sizes or establish causality
- Significant reporting lag (months to years)
- Subject to stimulated reporting and notoriety bias

Pre-Market Clinical Trials:

- Limited sample sizes (typically hundreds to low thousands)
- Short duration (weeks to months)
- Homogeneous populations (exclusion criteria eliminate comorbidities)
- Controlled conditions unlike real-world use
- Cannot detect rare or delayed adverse events
- Cost: Average Phase III trial costs $20M and takes 3+ years[79]

Post-Market Studies:

- Expensive and time-consuming
- Often industry-sponsored with potential conflicts
- Limited to specific questions rather than comprehensive monitoring

3.3 The Real-World Data Opportunity

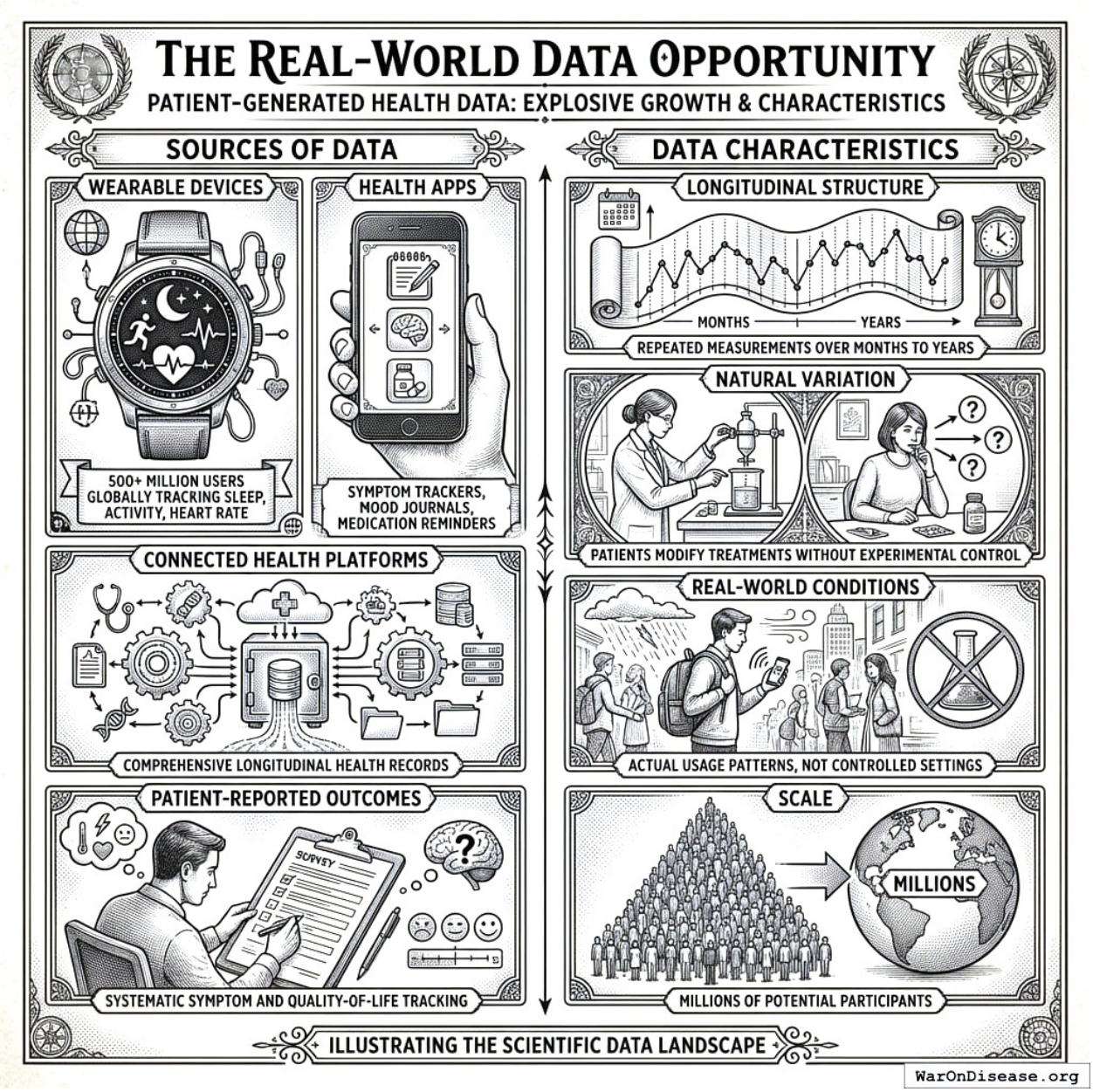

Figure 8: People voluntarily track their sleep, heart rate, mood, and bowel movements on their phones. You could use this to cure disease. You mostly use it to sell them running shoes.

The past decade has seen explosive growth in patient-generated health data:

- **Wearable devices**: 500+ million users globally tracking sleep, activity, heart rate[143]
- **Health apps**: Symptom trackers, mood journals, medication reminders
- **Connected health platforms**: Comprehensive longitudinal health records
- **Patient-reported outcomes**: Systematic symptom and quality-of-life tracking

This data is characterized by:

- **Longitudinal structure**: Repeated measurements over months to years
- **Natural variation**: Patients modify treatments without experimental control
- **Real-world conditions**: Actual usage patterns, not controlled settings
- **Scale**: Millions of potential participants

3.4 Our Contribution

We present a framework that transforms real-world health data into actionable pharmacovigilance intelligence:

1. **Quantitative Outcome Labels**: For each treatment, generate effect sizes (percent change from baseline) for all measured outcomes
2. **Treatment Rankings**: Rank treatments by efficacy and safety within therapeutic categories
3. **Automated Signal Detection**: Identify safety concerns (negative correlations) and efficacy signals (positive correlations)
4. **Bradford Hill Integration**: Composite scoring that operationalizes causal inference criteria[144,145]
5. **Scalable Implementation**: Analyze millions of treatment-outcome pairs automatically

This is not a replacement for RCTs but a complement, providing continuous, population-scale monitoring that can:

- Generate hypotheses for experimental validation
- Detect signals missed by spontaneous reporting
- Quantify effects that RCTs can only describe qualitatively
- Enable personalized benefit-risk assessment

Multiple meta-analyses demonstrate that well-designed observational studies produce effect sizes concordant with randomized controlled trials, supporting the validity of real-world evidence for hypothesis generation:

4 Data Collection and Integration

4.1 Data Sources

Our data integration protocol specifies how data flows from multiple sources, each contributing different variable types:

Source Category	Examples	Data Types
Wearables	Fitbit, Apple Watch, Oura Ring, Garmin	Sleep, steps, heart rate, HRV
Health Apps	Symptom trackers, mood journals	Symptoms, mood, energy, pain
Medication Trackers	Medisafe, MyTherapy	Drug intake, dosage, timing
Diet Trackers	MyFitnessPal, Cronometer	Foods, nutrients, calories
Lab Integrations	Quest, LabCorp APIs	Biomarkers, blood tests
EHR Connections	FHIR-enabled systems	Diagnoses, prescriptions, vitals
Manual Entry	Custom tracking	Any user-defined variable

Source Category	Examples	Data Types
Environmental	Weather APIs, air quality	Temperature, humidity, pollution

4.2 Variable Ontology

Variables are organized into semantic categories that inform default processing parameters:

Category	Examples	Onset Delay	Duration	Filling
Treatments	Drugs, supplements	30 min	24 hours	Zero
Foods	Diet, beverages	30 min	10 days	Zero
Symptoms	Pain, fatigue, nausea	0	24 hours	None
Emotions	Mood, anxiety, depression	0	24 hours	None
Vital Signs	Blood pressure, glucose	0	24 hours	None
Sleep	Duration, quality, latency	0	24 hours	None
Physical Activity	Steps, exercise, calories burned	0	24 hours	None
Environment	Weather, air quality, allergens	0	24 hours	None
Physique	Weight, body fat, measurements	0	7 days	None

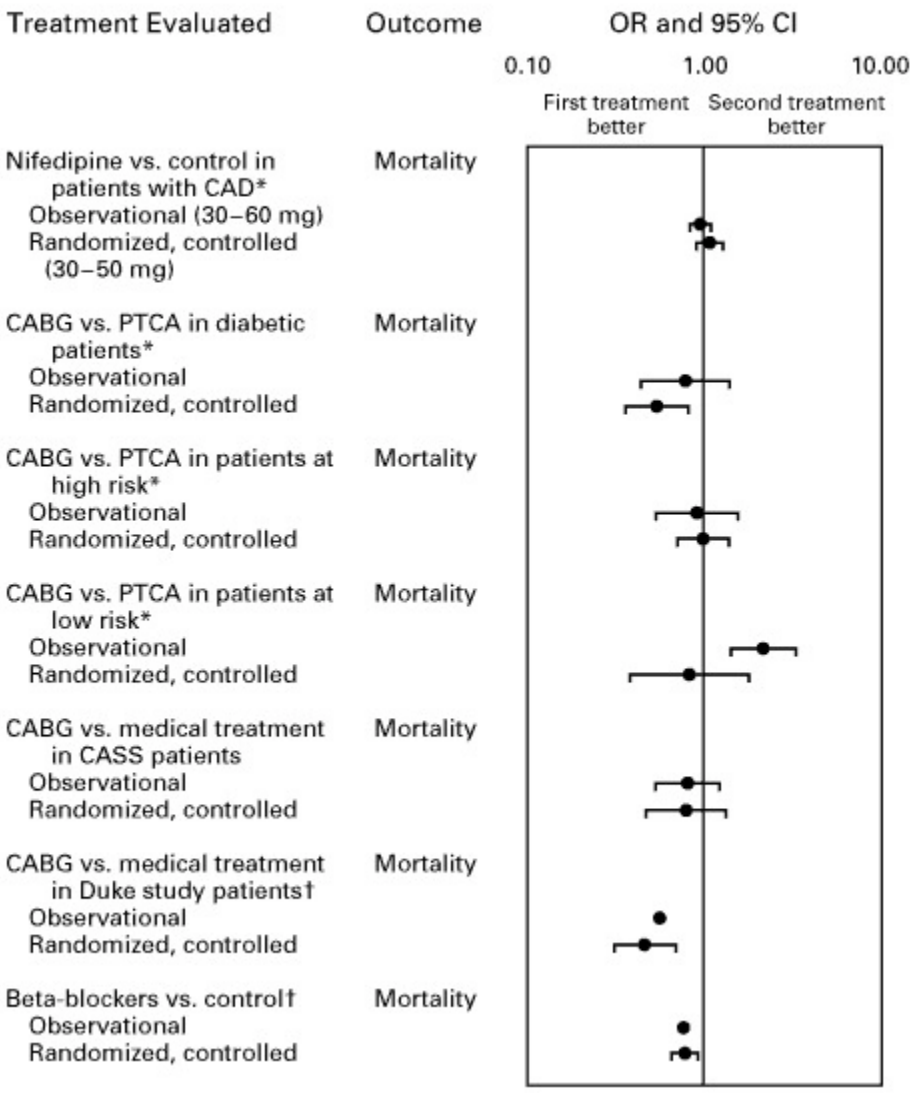

Figure 9: Turns out watching people die gives you the same answer as randomly choosing who dies. Science!

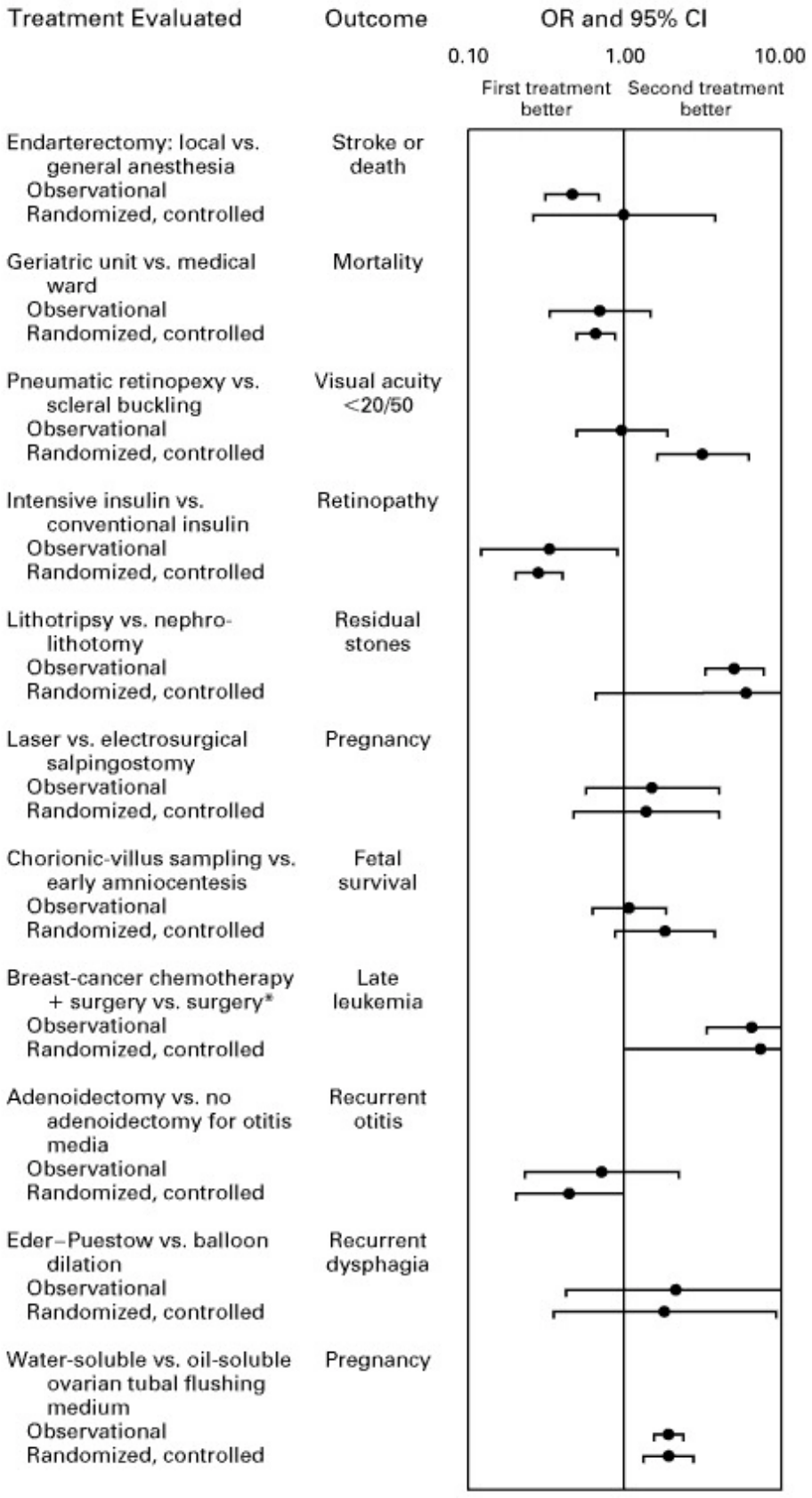

Figure 10: The fancy expensive experiments get the same results as just watching what happens. You've been overpaying for decades.

4.3 Measurement Structure

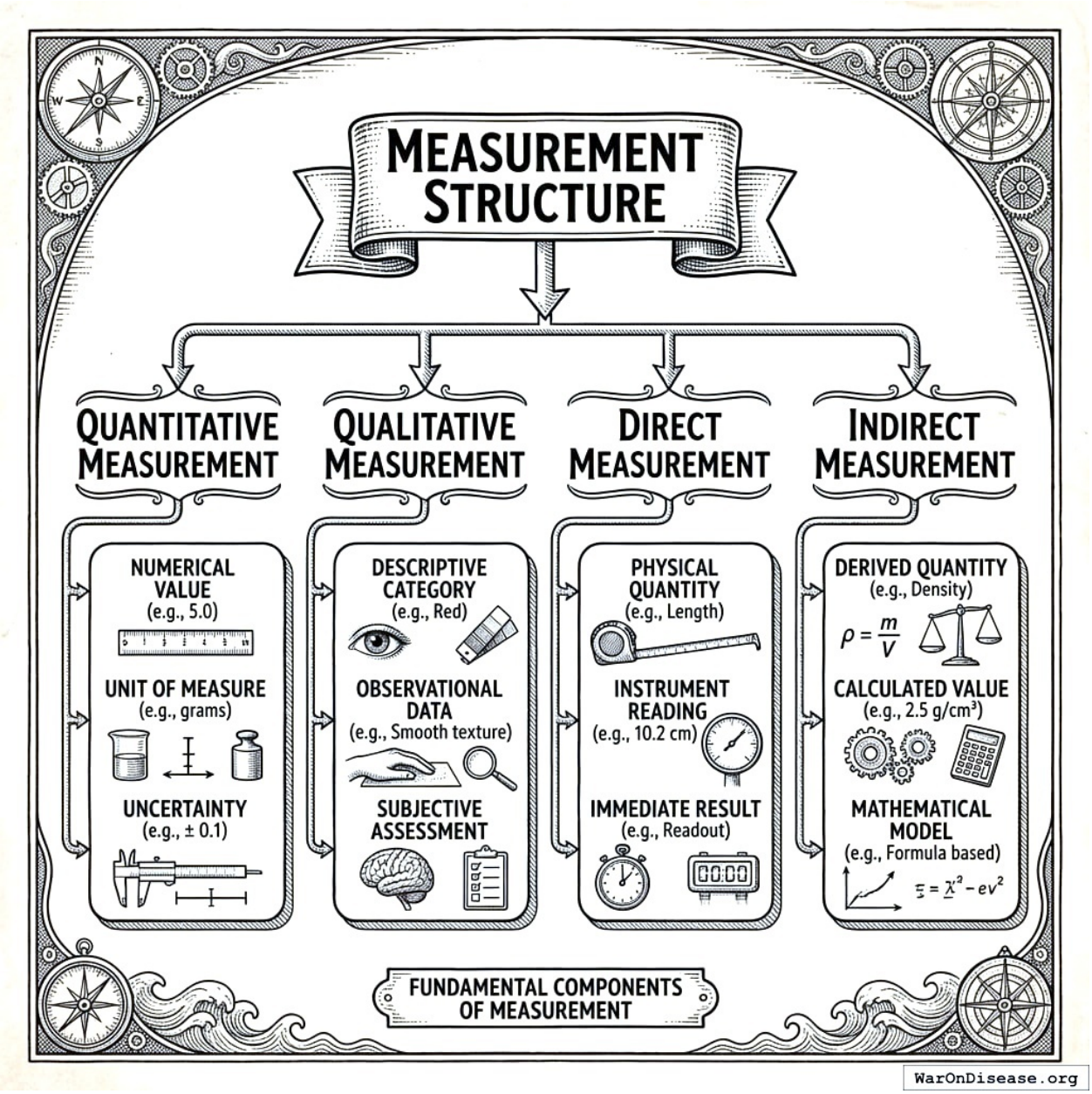

Figure 11: Every time you measure something, you have to write down who, what, when, and how. It's like a murder mystery, but for data points.

Each measurement includes:

```
Measurement {
    variable_id: int           // Reference to variable definition
    user_id: int               // Anonymized participant identifier
    value: float               // Numeric measurement value
    unit_id: int               // Standardized unit reference
    start_time: timestamp      // When measurement was taken
```

```
    source_id: int              // Data source for provenance
    note: string (optional)     // User annotation
}
```

4.4 Unit Standardization

The measurement standardization protocol converts all measurements to standardized units for cross-source compatibility:

- Weights → kilograms
- Distances → meters
- Temperatures → Celsius
- Dosages → milligrams
- Durations → seconds
- Percentages → 0-100 scale
- Ratings → 1-5 scale (normalized)

5 Mathematical Framework

The short version: We track what people take (treatments, supplements, foods) and how they feel (symptoms, mood, energy) over time. Then we look for patterns: "When people take more of X, does Y get better or worse?" We account for the fact that treatments take time to work (onset delay) and their effects fade (duration of action). The math below makes this rigorous.

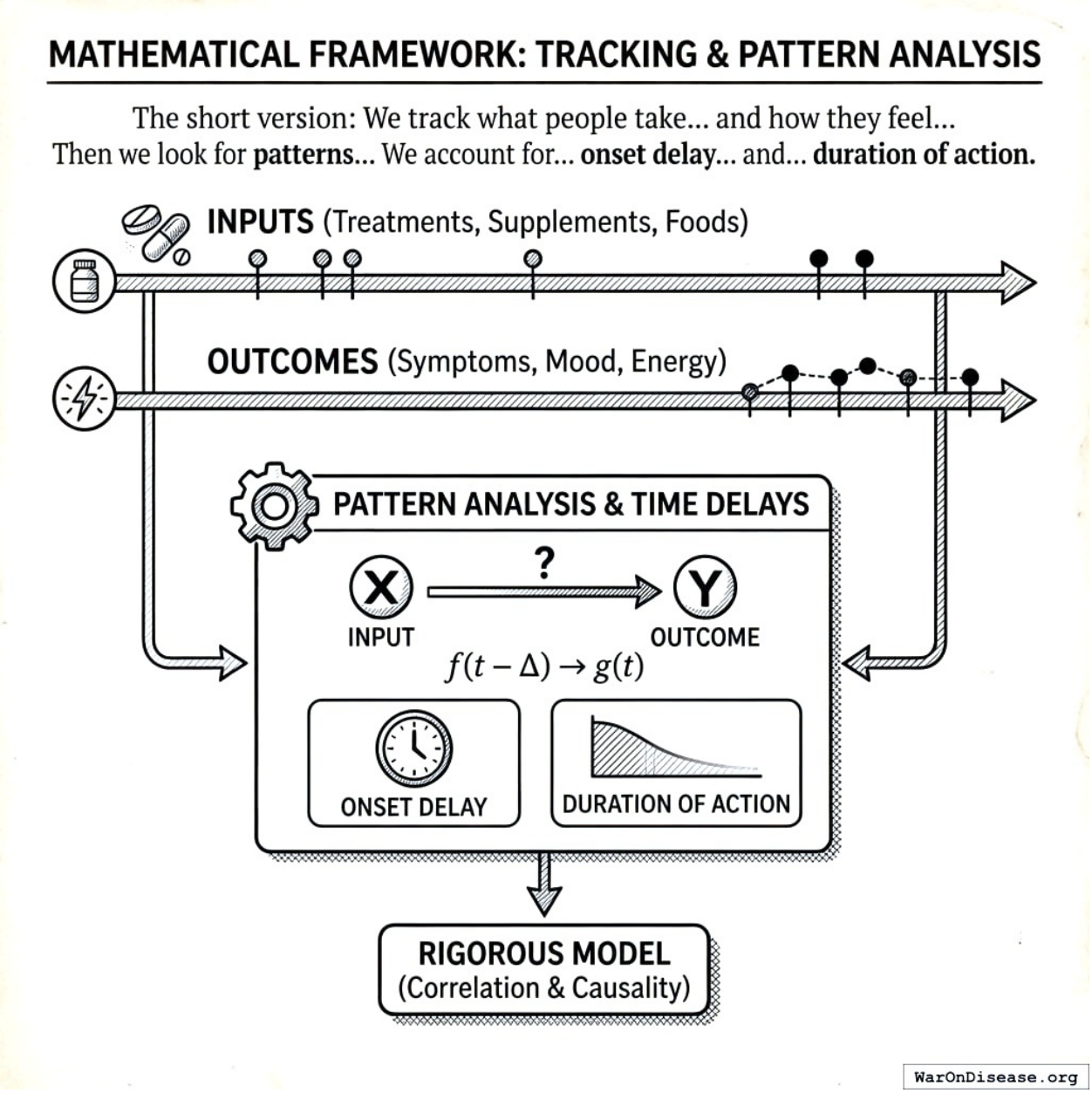

Figure 12: Take pill. Wait. Feel better. Feel worse again. Take another pill. You'd think medicine would have figured out the timing by now.

5.1 Data Structure

For each participant $i \in \{1, ..., N\}$, we observe time series of predictor variable P (e.g., treatment) and outcome variable O (e.g., symptom):

$$P_i = \{(t^P_{i,1}, p_{i,1}), (t^P_{i,2}, p_{i,2}), ..., (t^P_{i,n_i}, p_{i,n_i})\}$$

$$O_i = \{(t^O_{i,1}, o_{i,1}), (t^O_{i,2}, o_{i,2}), ..., (t^O_{i,m_i}, o_{i,m_i})\}$$

where t denotes timestamp, p denotes predictor measurements, and o denotes outcome measurements. Critically, timestamps need not be aligned. The temporal alignment protocol handles asynchronous, irregular sampling.

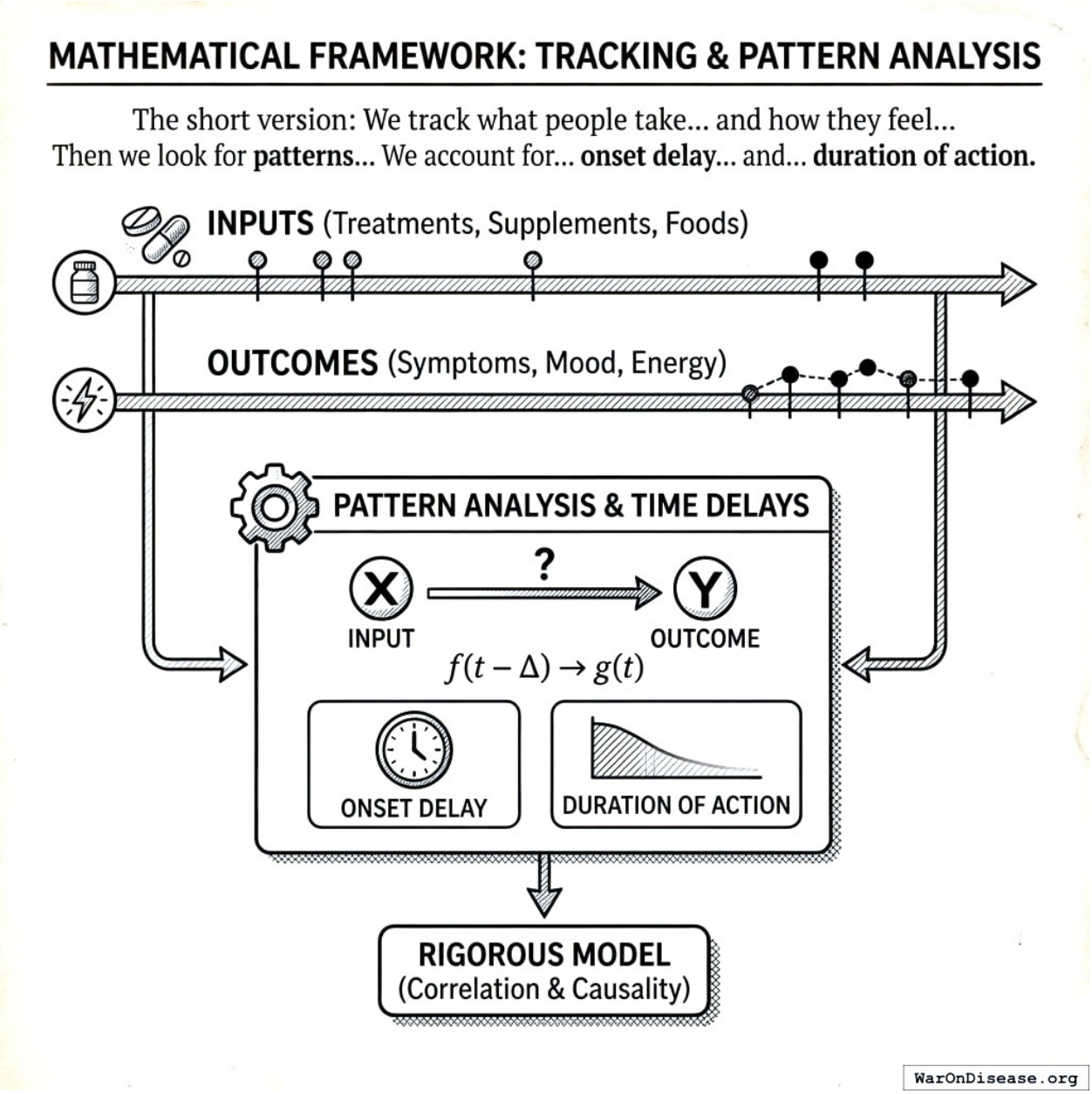

Figure 13: If you take aspirin at noon, your headache goes away around 12:30 and comes back at 5pm. Computers need Greek letters to understand this.

5.2 Temporal Alignment

5.2.1 Onset Delay and Duration of Action

Treatments do not produce immediate effects. We define:

- **Onset delay** δ: Time lag before treatment produces observable effect
- **Duration of action** τ: Time window over which effect persists

Constraints:

$$0 \leq \delta \leq 8{,}640{,}000 \text{ seconds (100 days)}$$

$$600 \leq \tau \leq 7{,}776{,}000 \text{ seconds (90 days)}$$

5.2.2 Outcome Window Calculation

For a predictor measurement at time t, we associate it with outcome measurements in the window:

$$W(t) = \{t_j : t + \delta \leq t_j \leq t + \delta + \tau\}$$

The aligned outcome value is computed as the mean:

$$\bar{o}(t) = \frac{1}{|W(t)|} \sum_{t_j \in W(t)} o_j$$

5.3 Pair Generation Strategies

We employ two complementary strategies depending on variable characteristics:

5.3.1 Outcome-Based Pairing (Predictor has Filling Value)

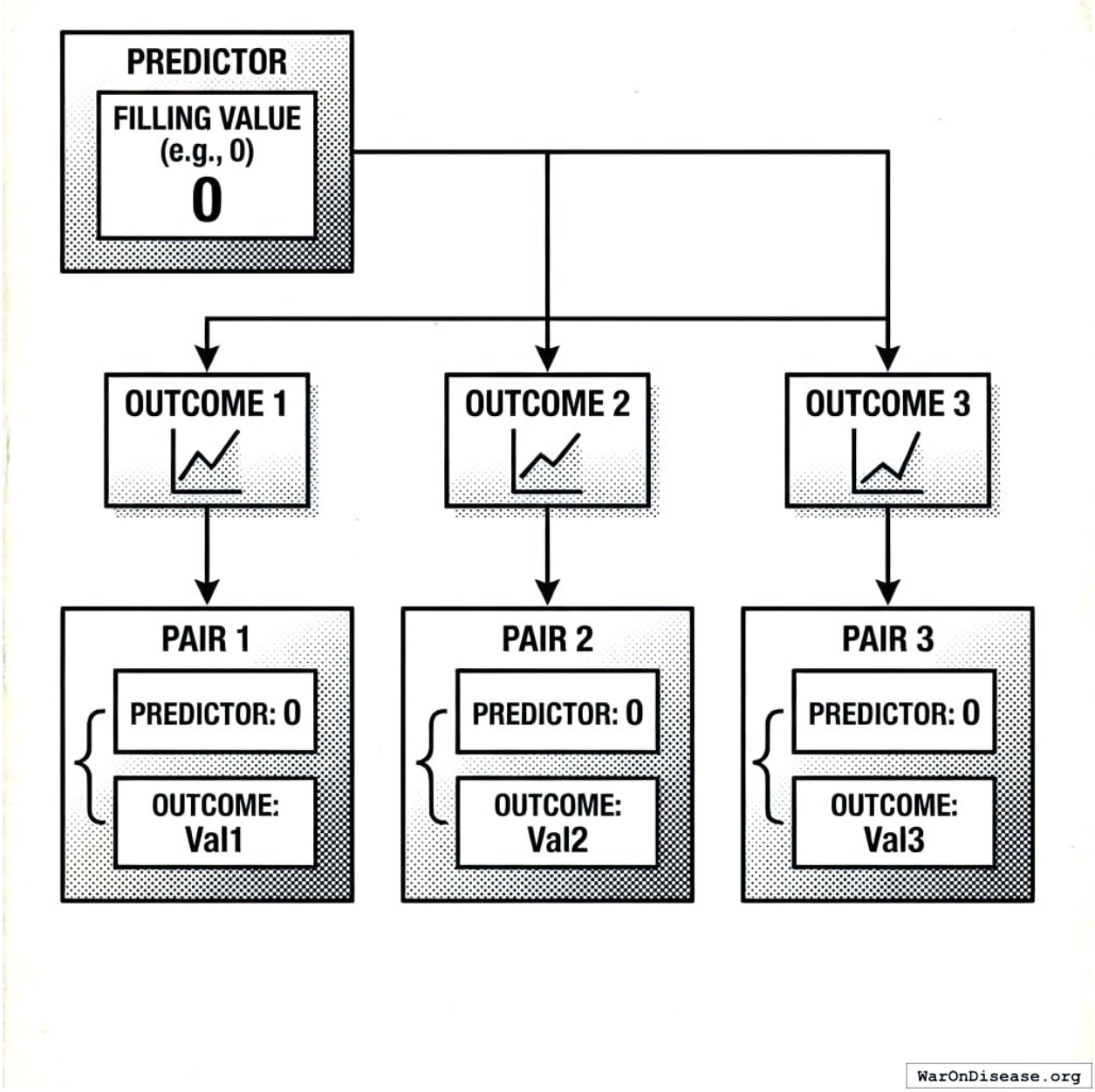

Figure 14: To know if the pill worked, you have to look backwards in time to see if you took it. Time travel, but boring.

When the predictor has a filling value (e.g., zero for "not taken"), we create one pair per outcome measurement:

```
For each outcome measurement (t_o, o):
    window_end = t_o - delta
    window_start = window_end - tau + 1
```

```
predictor_values = measurements in [window_start, window_end]

if predictor_values is empty:
    predictor_value = filling_value  // e.g., 0
else:
    predictor_value = mean(predictor_values)

create_pair(predictor_value, o)
```

5.3.2 Predictor-Based Pairing (No Filling Value)

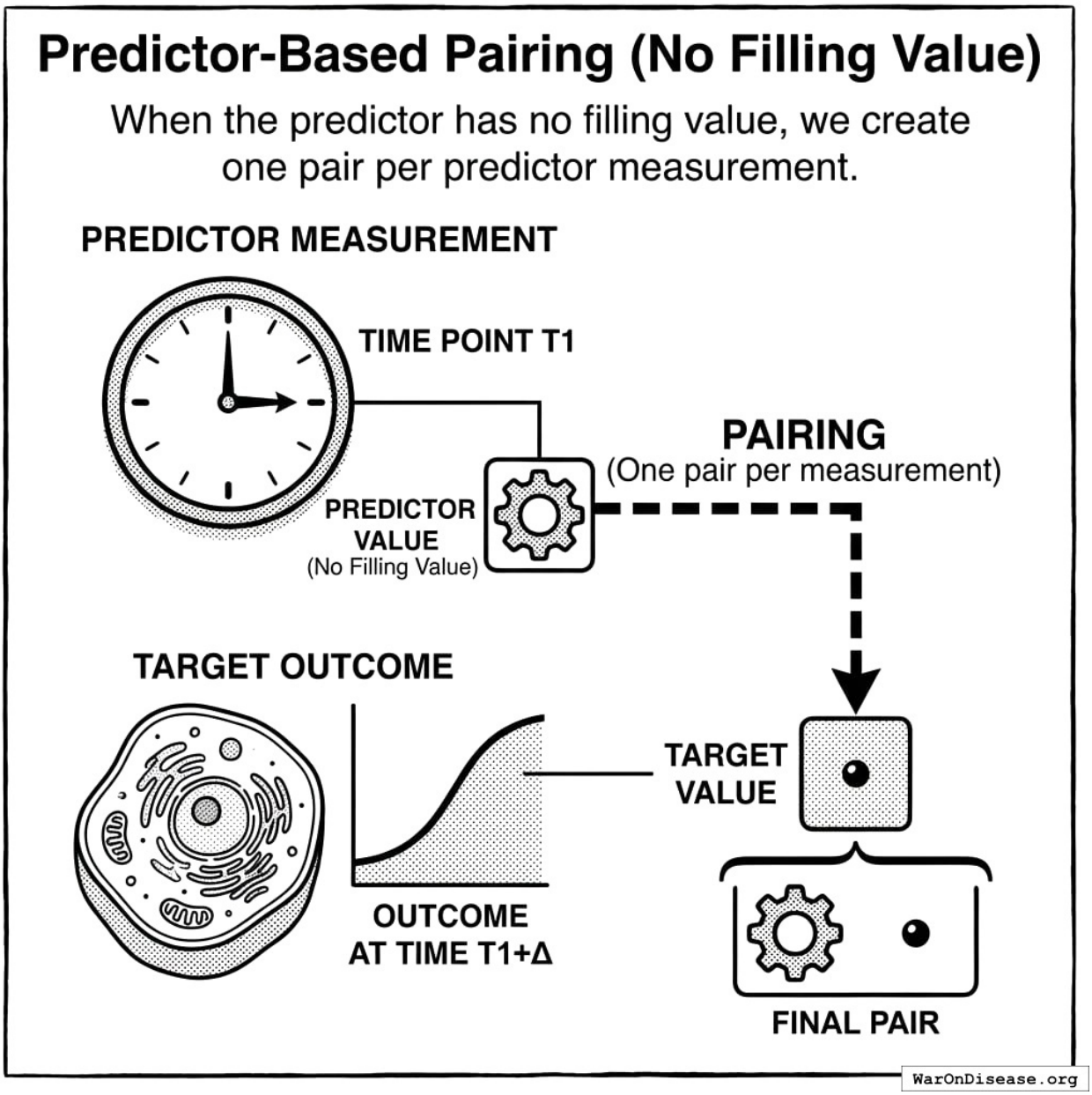

Figure 15: Medicine happens. Time passes. Body does things. You measure the things. It's called 'waiting' but scientists need diagrams.

When the predictor has no filling value, we create one pair per predictor measurement:

```
For each predictor measurement (t_p, p):
    window_start = t_p + delta
    window_end = window_start + tau - 1

    outcome_values = measurements in [window_start, window_end]
```

```
if outcome_values is empty:
    skip this pair
else:
    outcome_value = mean(outcome_values)
    create_pair(p, outcome_value)
```

5.4 Filling Value Logic

5.4.1 Filling Types

Type	Description	Use Case
Zero	Missing = 0	Treatments (assume not taken)
Value	Missing = specific constant	Known default states
None	No imputation	Continuous outcomes
Interpolation	Linear interpolation	Slowly-changing variables

5.4.2 Temporal Boundaries

To prevent spurious correlations from extended filling periods:

- **Earliest filling time**: First recorded measurement (tracking start)
- **Latest filling time**: Last recorded measurement (tracking end)

Pairs outside these boundaries are excluded. This prevents filling zeros for a treatment before the participant started tracking it.

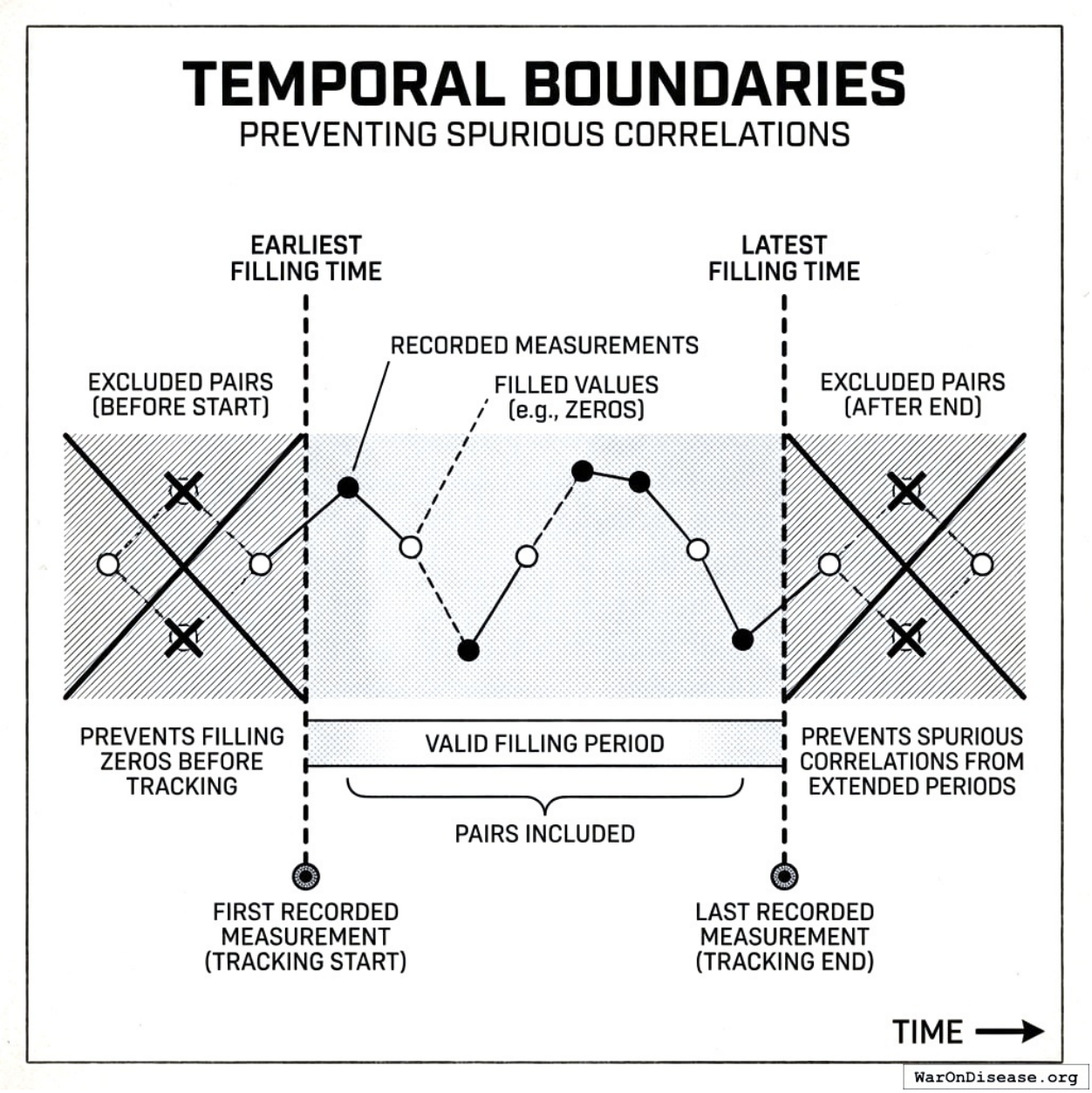

Figure 16: Only use data from when people were actually paying attention. Ignore measurements from that week they forgot their tracking app existed.

5.4.3 Conservative Bias

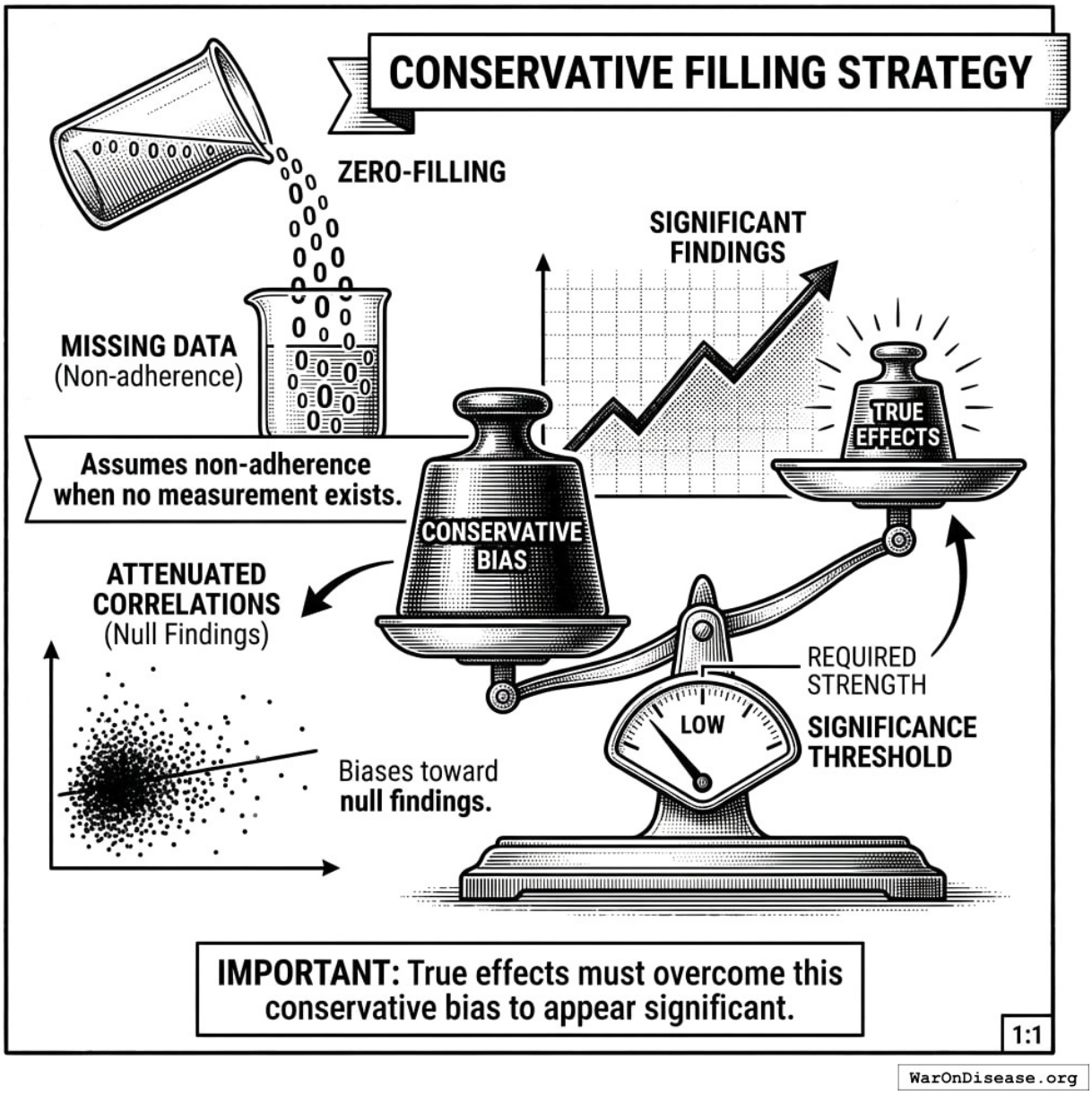

Figure 17: When people forget to log their data, pretend they took zero pills. This makes drugs look worse than they are, which is somehow the responsible thing to do.

Our filling strategy is deliberately conservative:

- Zero-filling for treatments assumes non-adherence when no measurement exists
- This biases toward null findings (attenuated correlations) rather than false positives
- True effects must overcome this conservative bias to appear significant

5.5 Baseline Definition and Outcome Estimation

5.5.1 Within-Subject Comparison

For each participant i, we compute the mean predictor value:

$$\bar{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} p_{i,j}$$

We partition measurements into **baseline** and **follow-up** periods:

$$\text{Baseline}_i = \{(p, o) : p < \bar{p}_i\}$$
$$\text{Follow-up}_i = \{(p, o) : p \geq \bar{p}_i\}$$

This creates a natural within-subject comparison:

- **Baseline**: Periods of below-average predictor exposure
- **Follow-up**: Periods of above-average predictor exposure

5.5.2 Outcome Means

$$\mu_{\text{baseline},i} = \mathbb{E}[o \mid p < \bar{p}_i]$$
$$\mu_{\text{follow-up},i} = \mathbb{E}[o \mid p \geq \bar{p}_i]$$

5.6 Percent Change from Baseline

The primary effect size metric:

$$\Delta_i = \frac{\mu_{\text{follow-up},i} - \mu_{\text{baseline},i}}{\mu_{\text{baseline},i}} \times 100$$

Advantages:

- **Interpretability**: "15% reduction in pain" is intuitive
- **Scale invariance**: Enables comparison across different outcome measures
- **Clinical relevance**: Standard metric in medical literature
- **Regulatory familiarity**: FDA uses percent change in efficacy assessments

5.7 Correlation Coefficients

We compute both parametric and non-parametric measures:

5.7.1 Pearson Correlation (Linear Relationships)

$$r_{\text{Pearson}} = \frac{\sum_{j=1}^{n}(p_j - \bar{p})(o_j - \bar{o})}{\sqrt{\sum_{j=1}^{n}(p_j - \bar{p})^2} \cdot \sqrt{\sum_{j=1}^{n}(o_j - \bar{o})^2}}$$

5.7.2 Spearman Rank Correlation (Monotonic Relationships)

$$r_{\text{Spearman}} = 1 - \frac{6 \sum_{j=1}^{n} d_j^2}{n(n^2 - 1)}$$

where $d_j = \text{rank}(p_j) - \text{rank}(o_j)$.

5.7.3 Forward and Reverse Correlations

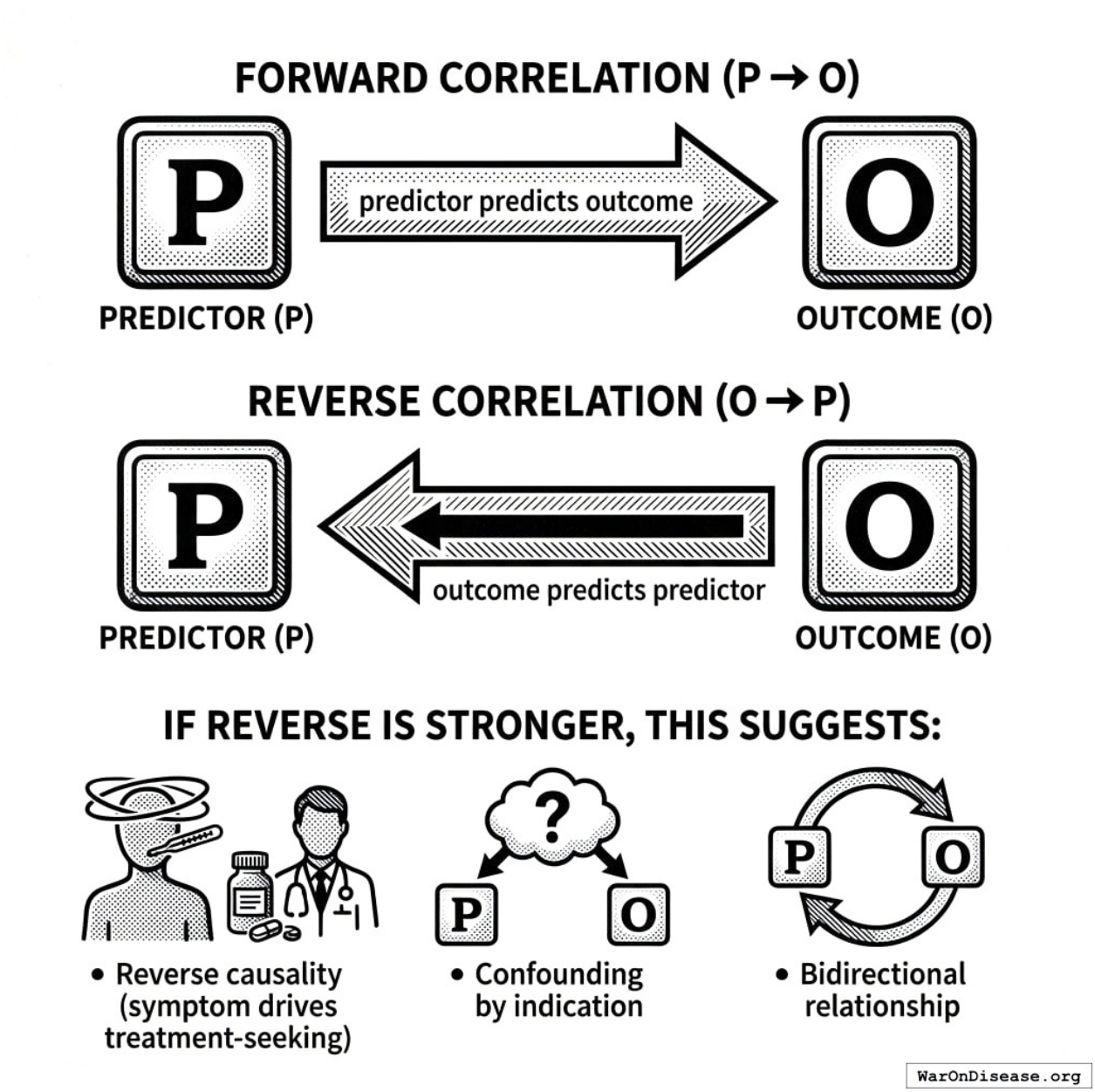

Figure 18: Does taking aspirin cure your headache, or does having a headache make you take aspirin? Computers get confused about which direction time flows.

We compute both:

- **Forward**: $P \to O$ (predictor predicts outcome)
- **Reverse**: $O \to P$ (outcome predicts predictor)

If reverse correlation is stronger, this suggests:

- Reverse causality (symptom drives treatment-seeking)
- Confounding by indication
- Bidirectional relationship

5.8 Z-Score Normalization

To assess effect magnitude relative to natural variability:

$$z = \frac{|\Delta|}{\text{RSD}_{\text{baseline}}}$$

where relative standard deviation:

$$\text{RSD}_{\text{baseline}} = \frac{\sigma_{\text{baseline}}}{\mu_{\text{baseline}}} \times 100$$

Interpretation: $z > 2$ indicates $p < 0.05$ under normality, meaning the observed effect exceeds typical baseline fluctuation.

5.9 Statistical Significance

Two-tailed t-test for correlation significance:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

with $n - 2$ degrees of freedom. Reject null hypothesis ($H_0 : \rho = 0$) at $\alpha = 0.05$ when:

$$|t| > t_{\text{critical}}(n-2, \alpha/2)$$

5.10 Hyperparameter Optimization

The onset delay δ^* and duration of action τ^* are selected to maximize correlation coefficient strength:

$$(\delta^*, \tau^*) = \underset{\delta,\tau}{\text{argmax}} \, |r(\delta, \tau)|$$

Search Strategy: 1. Initialize with category defaults (e.g., 30 min onset, 24 hr duration for drugs) 2. Grid search over physiologically plausible ranges 3. Select parameters yielding strongest correlation coefficient

Overfitting Mitigation:

- Restrict search to category-appropriate ranges
- Require minimum sample size before optimization
- Report both optimized and default-parameter results

6 Population Aggregation

6.1 Individual to Population

For population-level estimates, aggregate across N participants:

$$\bar{r} = \frac{1}{N} \sum_{i=1}^{N} r_i$$

$$\bar{\Delta} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i$$

6.2 Standard Error and Confidence Intervals

$$\text{SE}_{\bar{r}} = \frac{\sigma_r}{\sqrt{N}}$$

$$\text{CI}_{95\%} = \bar{r} \pm 1.96 \cdot \text{SE}_{\bar{r}}$$

6.3 Heterogeneity Assessment

Between-participant variance:

$$\sigma^2_{\text{between}} = \text{Var}(r_i)$$

High heterogeneity suggests:

- Subgroup effects (responders vs. non-responders)
- Interaction with unmeasured factors
- Need for personalized analysis

7 Data Quality Requirements

7.1 Minimum Thresholds

Requirement	Threshold	Rationale
Predictor value changes	≥ 5	Ensures sufficient variance
Outcome value changes	≥ 5	Ensures sufficient variance
Overlapping pairs	≥ 30	Central limit theorem
Baseline fraction	$\geq 10\%$	Adequate baseline
Follow-up fraction	$\geq 10\%$	Adequate predictor exposure
Processed daily measurements	≥ 4	Minimum data density

7.2 Variance Validation

Before computing variable relationships, validate sufficient variance:

$$\text{changes}(X) = \sum_{j=1}^{n-1} \mathbb{1}[x_j \neq x_{j+1}]$$

If $\text{changes}(P) < 5$ or $\text{changes}(O) < 5$, abort with `InsufficientVarianceException`.

7.3 Outcome Value Spread

$$\text{spread}_O = \max(O) - \min(O)$$

Variable relationships with zero spread are undefined and excluded.

8 Predictor Impact Score

The short version: Not all correlations are created equal. If we observe that "people who take Drug X report less pain," how confident should we be? The Predictor Impact Score (PIS) answers this by combining: (1) how strong is the relationship, (2) how many people show it, (3) does the drug come *before* the improvement (not after), and (4) is there a dose-response pattern. High PIS = worth investigating in a clinical trial. Low PIS = probably noise.

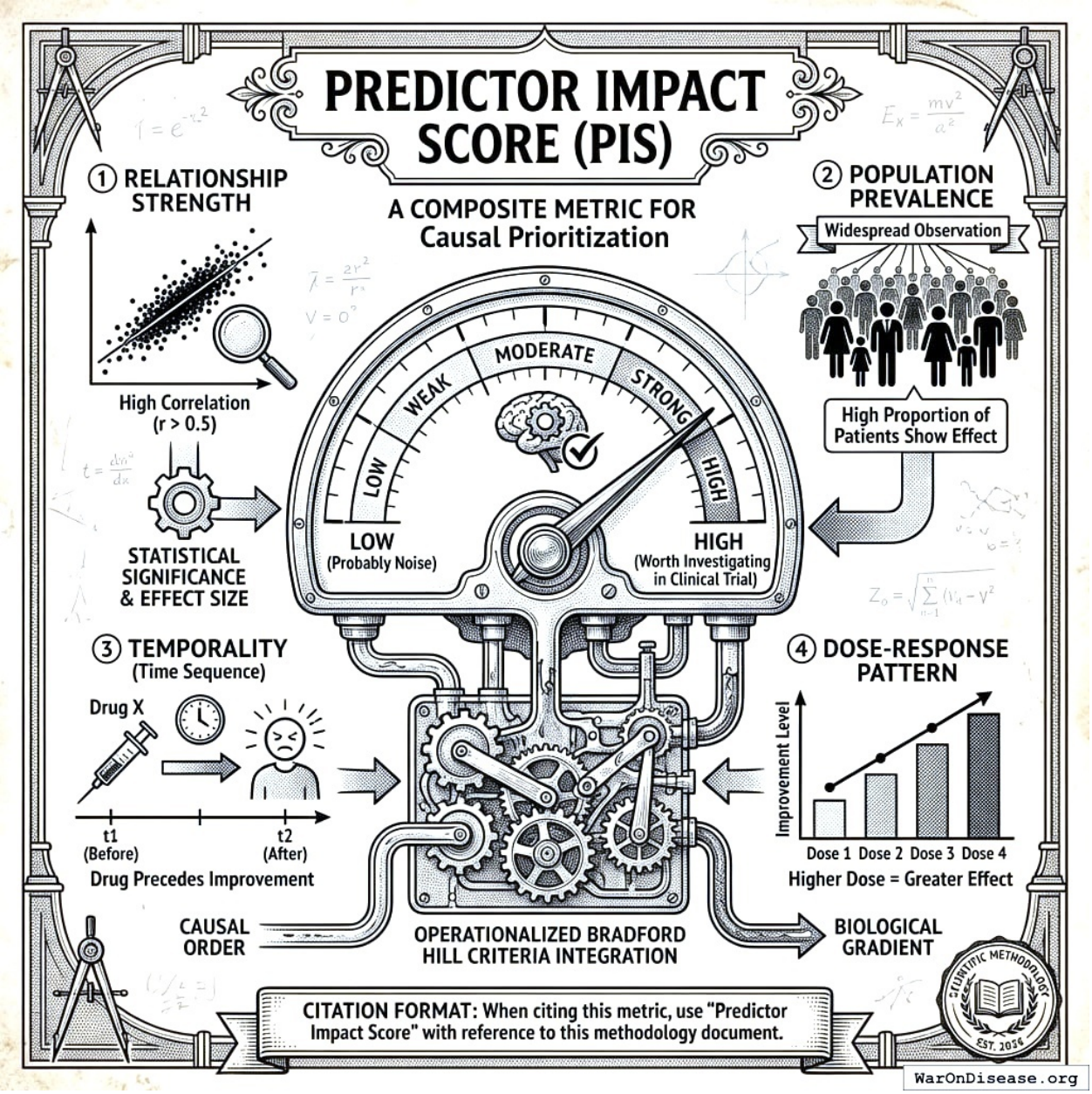

Figure 19: Four ways to tell if a drug actually works or if you're just seeing patterns in random noise, like Jesus in toast.

The **Predictor Impact Score (PIS)** is a composite metric that quantifies treatment-outcome relationship strength from patient health data, operationalizing Bradford Hill causality criteria to prioritize drug effects for clinical trial validation. It integrates correlation strength, statistical significance, effect magnitude, and multiple Bradford Hill criteria into a single interpretable score. Higher scores indicate predictors with greater, more reliable impact on the outcome.

> **Citation format**: When citing this metric in academic work, use "Predictor Impact Score" with reference to this methodology document.

8.1 What Makes the Predictor Impact Score Novel

Unlike simple correlation coefficients, PIS addresses fundamental limitations of observational analysis:

1. **Sample size agnosticism**: Raw correlations don't account for whether N=10 or N=10,000. PIS incorporates saturation functions that weight evidence accumulation.
2. **Temporal ambiguity**: Correlations can't distinguish A→B from B→A. PIS includes a temporality factor comparing forward vs. reverse correlations.
3. **Effect magnitude blindness**: Statistical significance practical significance. PIS incorporates z-scores to assess effect magnitude relative to baseline variability.
4. **Isolated metrics**: Traditional analysis reports correlation, p-value, and effect size separately. PIS integrates them into a single prioritization metric aligned with Bradford Hill criteria.

The Predictor Impact Score is not a causal proof. It's a principled heuristic for ranking which predictor-outcome relationships warrant further investigation, including experimental validation.

8.2 User-Level Predictor Impact Score

For individual participant (N-of-1) analyses, we compute:

$$\text{PIS}_{\text{user}} = |r| \cdot S \cdot \phi_z \cdot \phi_{\text{temporal}} \cdot f_{\text{interest}} + \text{PIS}_{\text{agg}}$$

Where:

- $|r|$ = absolute value of the correlation coefficient (strength)
- S = statistical significance (1 - p-value)
- ϕ_z = normalized z-score factor (effect magnitude)
- ϕ_{temporal} = temporality factor (forward vs. reverse causation)
- f_{interest} = interest factor (penalizes spurious variable pairs)
- PIS_{agg} = population-level aggregate score (provides context from broader population)

8.3 Aggregate (Population-Level) Predictor Impact Score

For population-level analyses aggregated across multiple participants:

$$\text{PIS}_{\text{agg}} = |r_{\text{forward}}| \cdot w \cdot \phi_{\text{users}} \cdot \phi_{\text{pairs}} \cdot \phi_{\text{change}} \cdot \phi_{\text{gradient}}$$

Where:

- $|r_{\text{forward}}|$ = absolute forward Pearson correlation coefficient (strength)
- w = weighted average of community votes on plausibility
- $\phi_{\text{users}} = 1 - e^{-N/N_{\text{sig}}}$ (user saturation, $N_{\text{sig}} = 10$)
- $\phi_{\text{pairs}} = 1 - e^{-n/n_{\text{sig}}}$ (pair saturation, n_{sig} = significant pairs threshold)
- $\phi_{\text{change}} = 1 - e^{-\Delta_{\text{spread}}/\Delta_{\text{sig}}}$ (change spread saturation)
- ϕ_{gradient} = biological gradient coefficient (dose-response)

The saturation functions asymptotically approach 1 as sample sizes increase, reflecting that consistent findings across more participants strengthen causal inference.

8.4 Z-Score and Effect Magnitude Factor

The z-score quantifies the magnitude of the outcome change relative to baseline variability:

$$z = \frac{|\Delta\%_{\text{baseline}}|}{\text{RSD}_{\text{baseline}}}$$

Where:

- $\Delta\%_{\text{baseline}}$ = percent change from baseline (see below)
- $\text{RSD}_{\text{baseline}}$ = relative standard deviation of outcome during baseline period

A z-score > 2 indicates statistical significance ($p < 0.05$), meaning the observed change is unlikely due to random variation.

The **normalized z-score factor** incorporates effect magnitude into the PIS score:

$$\phi_z = \frac{|z|}{|z| + z_{\text{ref}}}$$

Where $z_{\text{ref}} = 2$ (the conventional significance threshold). This saturating function:

- Approaches 0 for negligible effects ($z \to 0$)
- Equals 0.5 at the significance threshold ($z = 2$)
- Approaches 1 for very large effects ($z \to \infty$)

8.5 Temporality Factor

The temporality factor quantifies evidence that the predictor precedes and causes the outcome (rather than reverse causation):

$$\phi_{\text{temporal}} = \frac{|r_{\text{forward}}|}{|r_{\text{forward}}| + |r_{\text{reverse}}|}$$

Where:

- r_{forward} = correlation when predictor precedes outcome (P $\to$ O)
- r_{reverse} = correlation when outcome precedes predictor (O $\to$ P)

This factor:

- Equals 0.5 when forward and reverse correlations are equal (ambiguous causality)
- Approaches 1 when forward correlation dominates (supports predictor $\to$ outcome)
- Approaches 0 when reverse correlation dominates (suggests reverse causation or confounding by indication)

8.6 Percent Change from Baseline

The primary effect size metric expressing treatment impact:

$$\Delta\%_{\text{baseline}} = \frac{\bar{O}_{\text{follow-up}} - \bar{O}_{\text{baseline}}}{\bar{O}_{\text{baseline}}} \times 100$$

Where:

- $\bar{O}_{\text{follow-up}}$ = mean outcome value during follow-up period (after predictor exposure)
- $\bar{O}_{\text{baseline}}$ = mean outcome value during baseline period (before predictor exposure)

For outcomes measured in percentages or with zero baseline, we use absolute change instead:

$$\Delta_{\text{abs}} = \bar{O}_{\text{follow-up}} - \bar{O}_{\text{baseline}}$$

8.7 Statistical Significance

The statistical significance component captures confidence in the relationship:

$$S = 1 - p$$

Where p is the p-value from the correlation significance test. Higher values indicate greater confidence that the observed relationship is not due to chance.

8.8 Interest Factor

The interest factor f_{interest} penalizes likely spurious or uninteresting variable pairs:

$$f_{\text{interest}} = f_P \cdot f_O \cdot f_{\text{pair}}$$

Where:

- f_P = predictor interest factor (reduced for test variables, apps, addresses)
- f_O = outcome interest factor (reduced for non-outcome categories)
- f_{pair} = pair appropriateness (reduced for illogical category combinations)

8.9 Additional Data Quality Components

Skewness Coefficient (penalizes non-normal distributions):

$$\phi_{\text{skew}} = \frac{1}{1+\gamma_P^2} \cdot \frac{1}{1+\gamma_O^2}$$

Kurtosis Coefficient (penalizes heavy tails):

$$\phi_{\text{kurt}} = \frac{1}{1+\kappa_P^2} \cdot \frac{1}{1+\kappa_O^2}$$

Biological Gradient (dose-response relationship):

$$\phi_{\text{gradient}} = \left(\frac{\bar{p}_{\text{high}} - \bar{p}}{\sigma_P} - \frac{\bar{p}_{\text{low}} - \bar{p}}{\sigma_P} \right)^2$$

Measures the standardized difference between predictor values that predict high vs. low outcomes.

8.10 Bradford Hill Criteria Mapping

The PIS operationalizes six of the nine Bradford Hill criteria for causality[146]:

Component	Formula	Bradford Hill Criterion	In PIS Formula
$\|r\|$	Correlation magnitude	**Strength**	Yes (direct)
ϕ_z	Normalized z-score	**Strength** (effect magnitude)	Yes (user-level)
$\Delta\%$	Percent change from baseline	**Strength** (clinical significance)	Yes (via ϕ_z)
$\phi_{\text{users}}, \phi_{\text{pairs}}$	Sample saturation	**Consistency**	Yes (aggregate)
ϕ_{gradient}	Dose-response coefficient	**Biological Gradient**	Yes (aggregate)
w	Weighted community votes	**Plausibility**	Yes (aggregate)
f_{interest}	Category appropriateness	**Specificity**	Yes (user-level)
ϕ_{temporal}	Forward/reverse ratio	**Temporality**	Yes (user-level)
$\delta > 0$	Onset delay requirement	**Temporality**	Enforced in design

8.11 Interpreting Predictor Impact Scores

9 Provisional Thresholds - Not Yet Validated

The PIS thresholds below are **theoretically motivated heuristics**, not empirically validated cutoffs. Until retrospective validation against RCT outcomes is performed (see Section 19), these thresholds should be treated as provisional guidelines for prioritization, not evidence standards.

PIS scores range from 0 to approximately 1 (though values slightly above 1 are possible with very strong evidence). Guidelines for interpretation:

PIS Range	Interpretation	Recommended Action
0.5	Strong evidence	High priority for RCT validation
0.3 - 0.5	Moderate evidence	Consider for experimental investigation
0.1 - 0.3	Weak evidence	Monitor for additional data
< 0.1	Insufficient evidence	Low priority; may be noise

Important caveats:

- These thresholds are preliminary and should be validated against RCT outcomes

- PIS is relative, not absolute. Use it for prioritization, not proof.
- High PIS does not guarantee causation; low PIS does not rule it out
- Context matters: a PIS of 0.2 for a novel relationship may be more interesting than 0.5 for a known one

9.1 Optimal Daily Value for Precision Dosing

A key output of our analysis is the **optimal daily value**, the predictor value that historically precedes the best outcomes. This enables personalized, precision dosing recommendations.

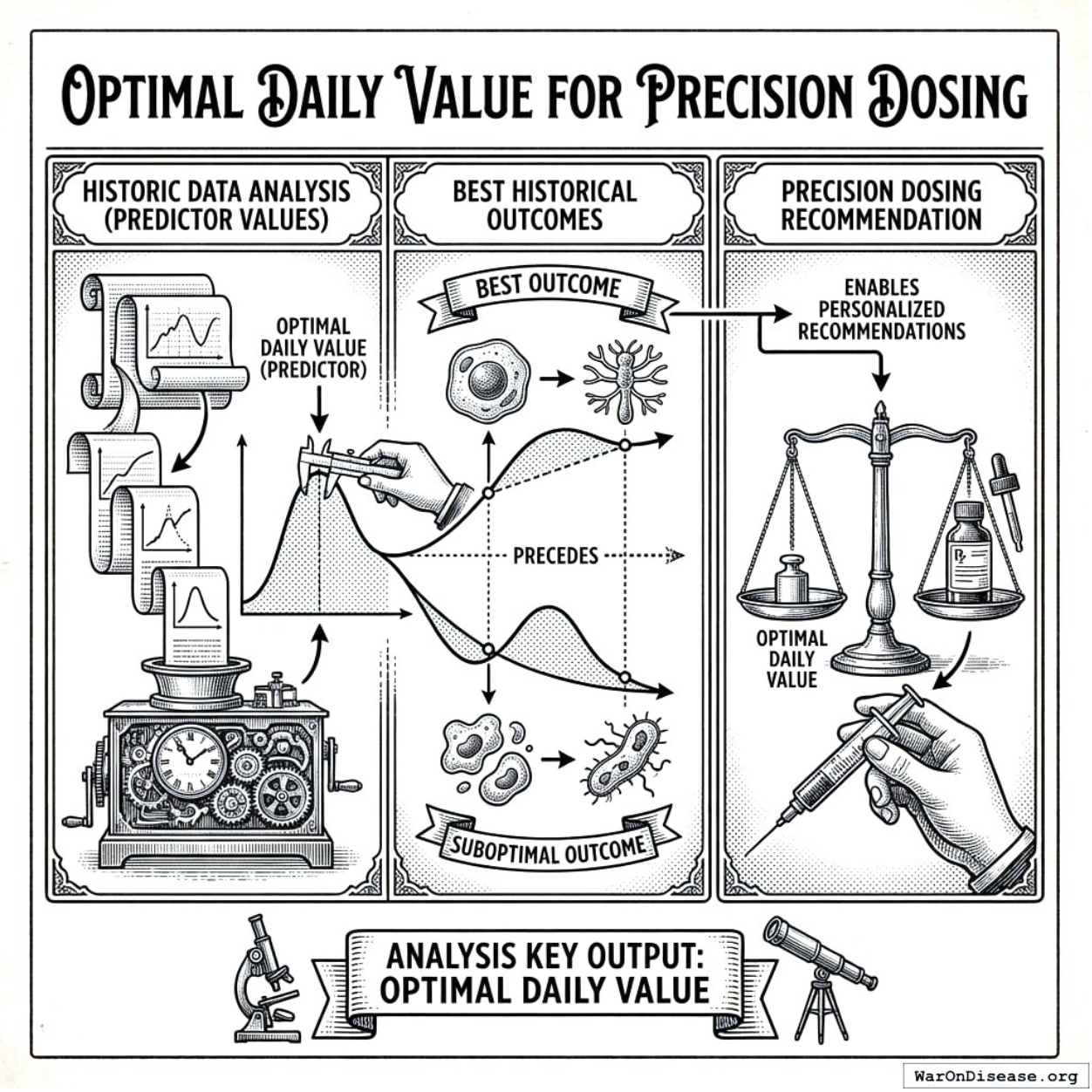

Figure 20: Computers look at what dose worked best for people like you in the past. It's astrology, but with math that actually works.

9.1.1 Value Predicting High Outcome

The **Value Predicting High Outcome** (V_{high}) is the average predictor value observed when the outcome exceeds its mean:

$$V_{\text{high}} = \frac{1}{|H|} \sum_{(p,o) \in H} p$$

Where:

- $H = \{(p, o) : o > \bar{O}\}$ is the set of predictor-outcome pairs where outcome exceeds its average
- $\bar{O}$ = mean outcome value across all pairs
- p = predictor (cause) value for each pair

Calculation Process: 1. Compute the average outcome value ($\bar{O}$) across all predictor-outcome pairs 2. Filter pairs to include only those where outcome $> \bar{O}$ (the "high effect" pairs) 3. Calculate the mean predictor value across these high-effect pairs

9.1.2 Value Predicting Low Outcome

The **Value Predicting Low Outcome** (V_{low}) is the average predictor value observed when the outcome is below its mean:

$$V_{\text{low}} = \frac{1}{|L|} \sum_{(p,o) \in L} p$$

Where:

- $L = \{(p, o) : o < \bar{O}\}$ is the set of predictor-outcome pairs where outcome is below its average

9.1.3 Grouped Optimal Values

For interpretability, we also calculate **grouped optimal values** that map to common dosing intervals:

- **Grouped Value Predicting High Outcome**: The nearest grouped predictor value (e.g., rounded to typical dosing units) to V_{high}
- **Grouped Value Predicting Low Outcome**: The nearest grouped predictor value to V_{low}

This allows recommendations like "400mg of Magnesium" rather than "412.7mg of Magnesium."

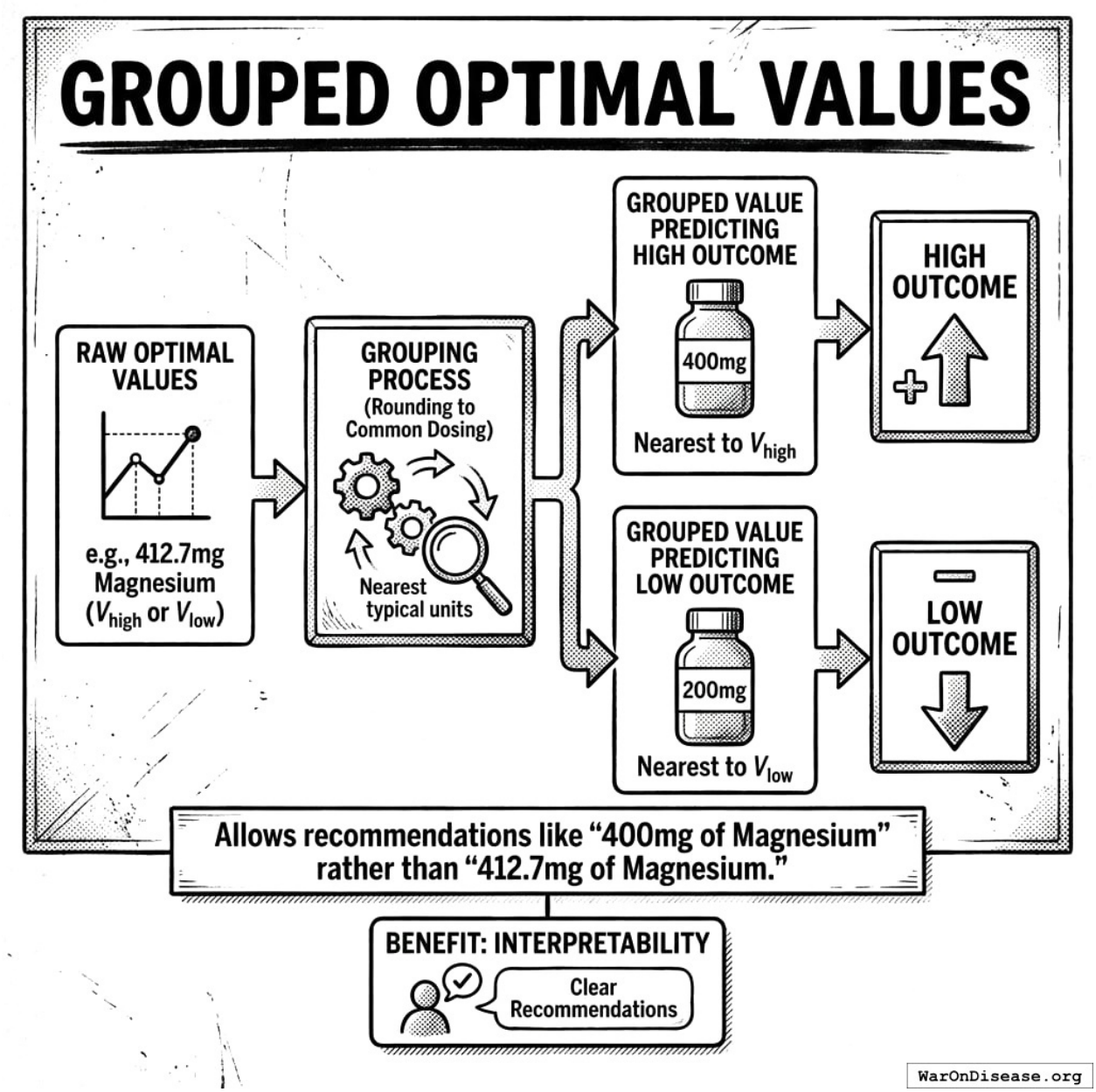

Figure 21: The computer says take 47.3mg. Your pills come in 50mg. Close enough, the computer sighs.

9.1.4 Precision Dosing Recommendations

These optimal values enable personalized recommendations:

For Positive Valence Outcomes (where higher is better, e.g., energy, sleep quality): > "Your [Outcome] was highest after [Grouped Value Predicting High Outcome] of [Predictor] over the previous [Duration of Action]." > > Example: "Your Sleep Quality was highest after 400mg of Magnesium over the previous 24 hours."

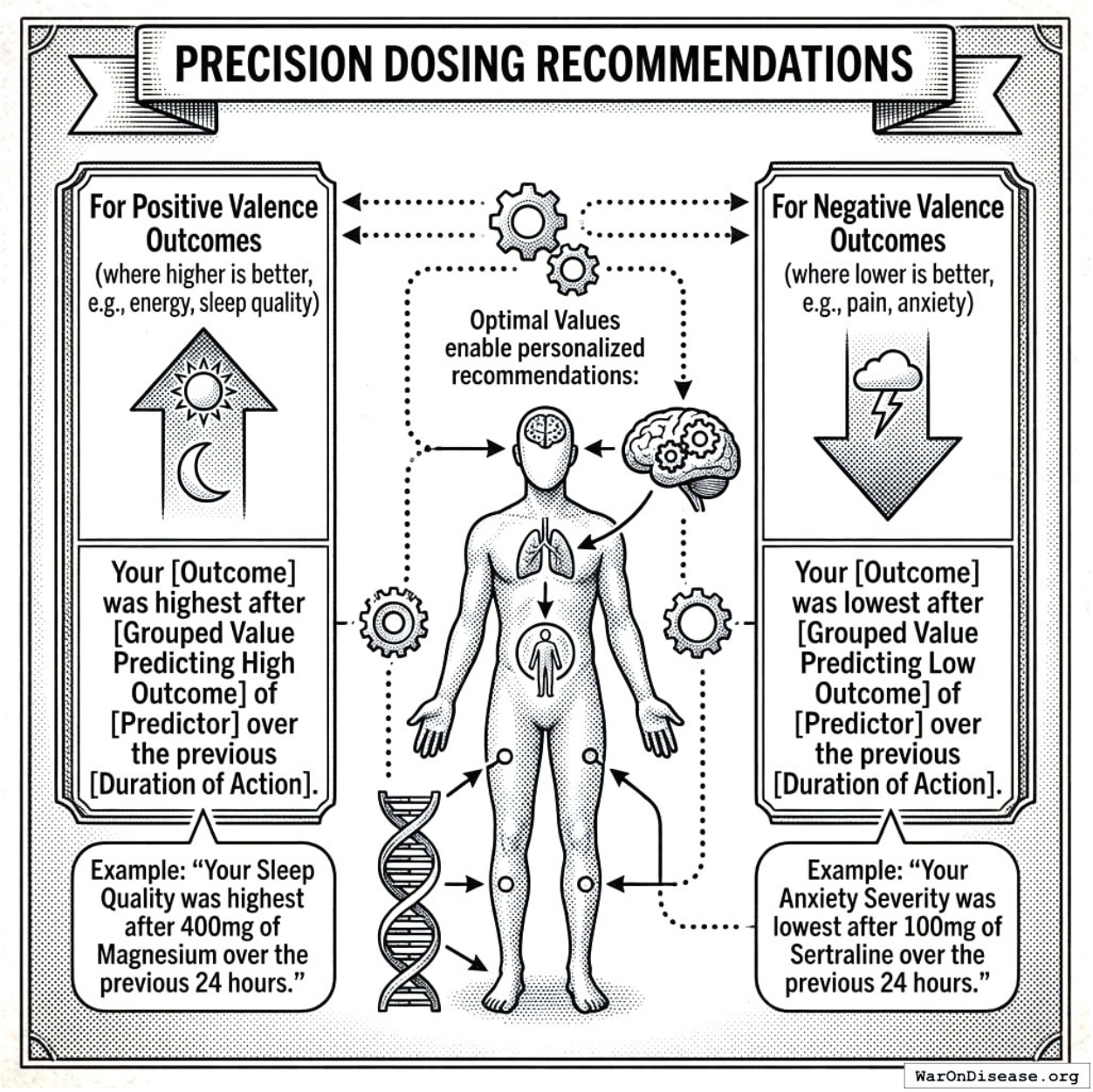

Figure 22: If more is better, take more. If more is worse, take less. You needed a flowchart for this.

For Negative Valence Outcomes (where lower is better, e.g., pain, anxiety): > "Your [Outcome] was lowest after [Grouped Value Predicting Low Outcome] of [Predictor] over the previous [Duration of Action]." > > Example: "Your Anxiety Severity was lowest after 100mg of Sertraline over the previous 24 hours."

9.1.5 Mathematical Relationship to Biological Gradient

The optimal values are closely related to the **biological gradient coefficient** (ϕ_{gradient}):

$$\phi_{\text{gradient}} = \left(\frac{V_{\text{high}} - \bar{P}}{\sigma_P} - \frac{V_{\text{low}} - \bar{P}}{\sigma_P} \right)^2$$

A larger separation between V_{high} and V_{low} indicates:

- Stronger dose-response relationship
- More reliable precision dosing recommendations
- Higher biological gradient coefficient

9.1.6 Clinical Applications

Metric	Definition	Clinical Use
V_{high}	Avg predictor when outcome > mean	Optimal dose for positive outcomes
V_{low}	Avg predictor when outcome < mean	Dose to avoid for positive outcomes
$V_{\text{high}} - V_{\text{low}}$	Optimal value spread	Magnitude of dose-response effect

Example Application: For a participant tracking Magnesium supplementation and Sleep Quality:

- $V_{\text{high}} = 412\text{mg} \rightarrow$ Grouped = 400mg (sleep quality highest after this dose)
- $V_{\text{low}} = 127\text{mg} \rightarrow$ Grouped = 125mg (sleep quality lowest after this dose)
- Recommendation: "Take approximately 400mg of Magnesium for optimal sleep quality"

9.1.7 Limitations

1. **Correlation Causation**: Optimal values reflect associations, not guaranteed causal effects
2. **Individual Variation**: Population optimal values may not be optimal for all individuals
3. **Context Dependence**: Optimal values may vary by timing, combination with other factors
4. **Grouping Artifacts**: Rounding to common doses may lose precision

Best Practice: Use optimal values as starting points for personal experimentation, not as definitive prescriptions.

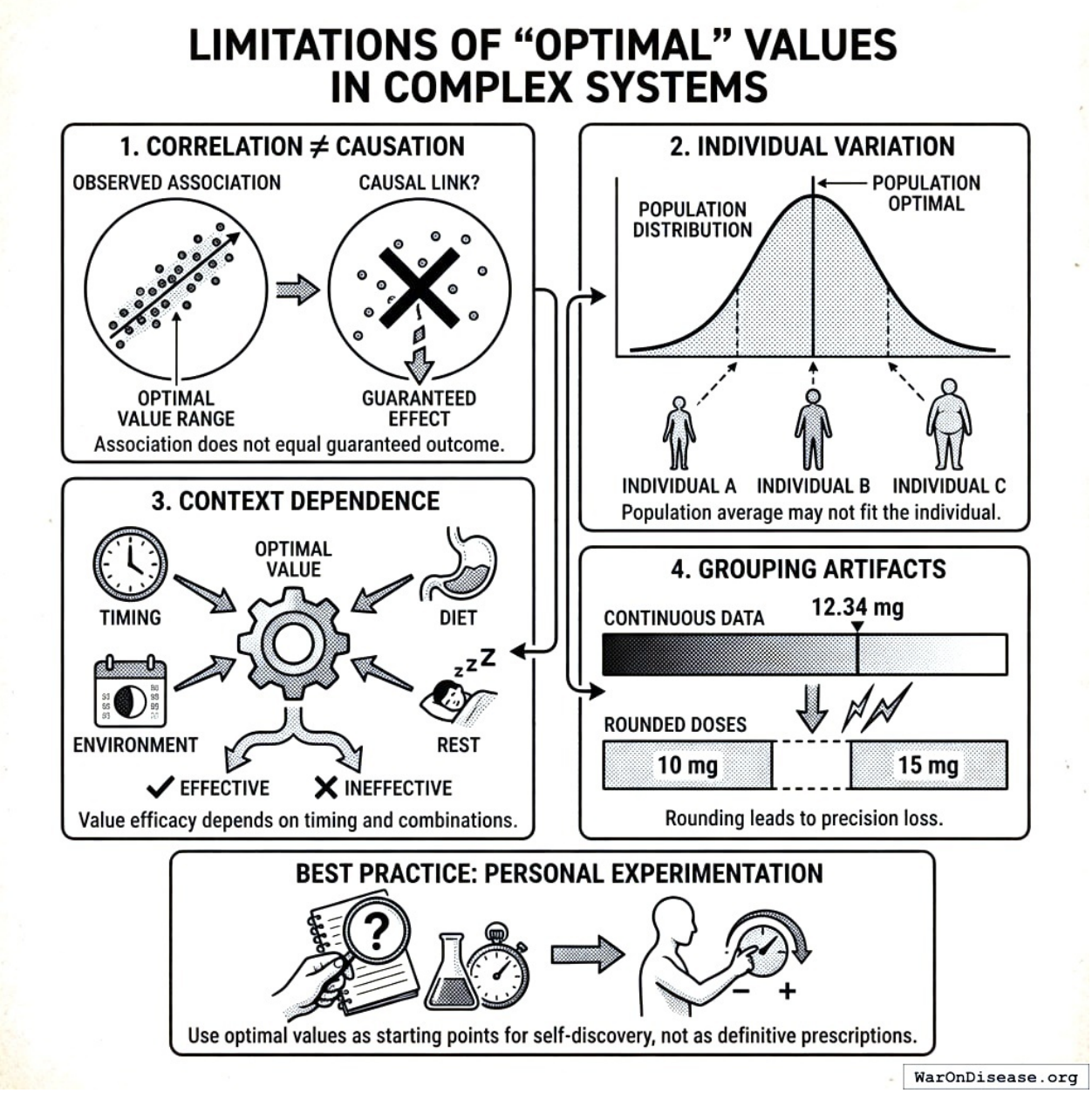

Figure 23: What works for most people is a starting point for figuring out what works for you specifically. Personalized medicine is just trial and error with better record keeping.

9.1.8 Confidence Intervals for Optimal Values

Optimal values should be reported with uncertainty bounds to convey reliability:

$$\text{CI}_{V_{\text{high}}} = V_{\text{high}} \pm t_{\alpha/2} \cdot \frac{\sigma_{p|H}}{\sqrt{|H|}}$$

Where:

- $\sigma_{p|H}$ = standard deviation of predictor values in high-outcome set H

- $|H|$ = number of pairs in high-outcome set
- $t_{\alpha/2}$ = critical t-value for desired confidence level

Interpretation Guidelines:

CI Width (relative to mean)	Reliability	Recommendation
$< 10\%$	High	Use as primary recommendation
10-25%	Moderate	Present as range (e.g., "350-450mg")
25-50%	Low	Insufficient precision for dosing
$> 50\%$	Very Low	Do not use for recommendations

Example: If $V_{\text{high}} = 400$mg with 95% CI [380, 420], report: "Optimal dose: 400mg (95% CI: 380-420mg)"

9.1.9 Individual vs Population Optimal Values

Both individual and population optimal values are computed and stored. Guidelines for use:

Scenario	Recommended Source	Rationale
User has 50 paired measurements	Individual V_{high}	Sufficient personal data
User has 20-50 measurements	Weighted blend	$0.5 \cdot V_{\text{user}} + 0.5 \cdot V_{\text{pop}}$
User has <20 measurements	Population V_{high}	Insufficient personal data
User's optimal differs >50% from population	Flag for review	May indicate unique response or data quality issue

Blending Formula:

$$V_{\text{recommended}} = w \cdot V_{\text{user}} + (1 - w) \cdot V_{\text{pop}}$$

Where $w = \min(1, n_{\text{user}}/n_{\text{threshold}})$ with $n_{\text{threshold}} = 50$ pairs.

9.1.10 Temporal Stability and Recalculation

Optimal values may drift over time due to:

- Physiological changes (age, weight, health status)
- Tolerance development
- Seasonal factors
- Lifestyle changes

Recalculation Policy:

Trigger	Action
New measurements added	Recalculate after every 10 new pairs
Time elapsed	Recalculate monthly regardless of new data
Significant life change	User-triggered recalculation
Optimal value drift >20%	Alert user to potential change

Rolling Window Option: For treatments where tolerance is expected, compute optimal values using only the most recent 90 days of data rather than all historical data.

Stability Metric:

$$\text{Stability} = 1 - \frac{|V_{\text{high}}^{\text{current}} - V_{\text{high}}^{\text{previous}}|}{V_{\text{high}}^{\text{previous}}}$$

Stability < 0.8 (>20% change) triggers a notification to the user.

9.1.11 Edge Cases: Minimal Dose-Response

When $V_{\text{high}} \approx V_{\text{low}}$, the predictor shows no clear dose-response relationship:

Detection Criterion:

$$\frac{|V_{\text{high}} - V_{\text{low}}|}{\sigma_P} < 0.5$$

(Less than half a standard deviation apart)

Possible Interpretations: 1. **Threshold effect**: Any dose above zero works equally well 2. **No effect**: Predictor doesn't influence outcome 3. **Non-linear response**: U-shaped or inverted-U curve not captured by simple high/low split 4. **Insufficient variance**: User takes similar doses, preventing detection

Handling:

- Do not display optimal value recommendations when dose-response is minimal
- Instead report: "No clear dose-response relationship detected for [Predictor] → [Outcome]"
- Flag for potential non-linear analysis in future versions

9.1.12 Validation of Optimal Values

The Critical Question: Do users who follow optimal value recommendations actually experience better outcomes than those who don't?

Proposed Validation Study:

1. **Prospective A/B Test**:
   - Group A: Receives personalized optimal value recommendations
   - Group B: Receives no recommendations (continues current behavior)
   - Compare outcome trajectories over 30-90 days
2. **Retrospective Adherence Analysis**:
   - For users with established optimal values, calculate "adherence score":

$$\text{Adherence} = \frac{\text{Days within } \pm 20\% \text{ of } V_{\text{high}}}{\text{Total tracking days}}$$

- Correlate adherence with outcome improvement

Success Metrics:

- Users in top adherence quartile should show >15% better outcomes than bottom quartile
- Optimal value recommendations should outperform random dosing by >10%

Current Status: This validation has not been performed. Until validated, optimal values should be presented as "data-driven suggestions" rather than "clinically validated recommendations."

9.2 Saturation Constant Rationale

The saturation constants (N_sig, n_sig, etc.) reflect pragmatic thresholds based on statistical and clinical considerations:

Constant	Value	Rationale
N_sig (users)	10	At N=10, user saturation 0.63. By N=30, 0.95. Reflects that consistency across 10+ individuals provides meaningful replication.
n_sig (pairs)	100	Central limit theorem suggests n 30 for normality. We use 100 as the "strong evidence" threshold.
Δ_sig (change spread)	10%	A 10% change is often considered clinically meaningful across many health outcomes.
z_ref	2	Corresponds to $p < 0.05$ under normality (the conventional significance threshold).

These constants are not empirically optimized. Future work should: 1. Validate constants against known causal relationships (from RCTs) 2. Consider domain-specific thresholds (e.g., psychiatric vs. cardiovascular outcomes) 3. Implement sensitivity analyses to assess robustness to constant choices

9.3 Effect Following High vs Low Predictor Values

Beyond optimal values, we calculate the **average outcome following different predictor levels** to quantify dose-response relationships:

9.3.1 Average Outcome Metrics

Metric	Definition	Clinical Interpretation
Average Outcome	Mean outcome across all pairs	Baseline outcome level

Metric	Definition	Clinical Interpretation
Average Outcome Following High Predictor	Mean outcome when predictor > mean	Outcome after high exposure
Average Outcome Following Low Predictor	Mean outcome when predictor < mean	Outcome after low exposure
Average Daily High Predictor	Mean predictor in upper 51% of spread	"High dose" value
Average Daily Low Predictor	Mean predictor in lower 49% of spread	"Low dose" value

9.3.2 Calculation

$$\bar{O}_{\text{high}} = \mathbb{E}[O \mid P > \bar{P}]$$
$$\bar{O}_{\text{low}} = \mathbb{E}[O \mid P \leq \bar{P}]$$

Where $\bar{P}$ is the mean predictor value across all pairs.

Effect Size from High to Low Cause:

$$\Delta_{\text{high-low}} = \frac{\bar{O}_{\text{high}} - \bar{O}_{\text{low}}}{\bar{O}_{\text{low}}} \times 100$$

This metric directly shows the percent difference in outcome between high and low predictor exposure periods.

9.4 Predictor Baseline and Treatment Averages

For treatment-response analysis, we distinguish between **baseline** (non-treatment) and **treatment** periods:

Metric	Definition	Use Case
Predictor Baseline Average Per Day	Average daily predictor during low-exposure periods	Typical non-treatment value
Predictor Treatment Average Per Day	Average daily predictor during high-exposure periods	Typical treatment dosage
Predictor Baseline Average Per Duration Of Action	Baseline cumulative over duration of action	For longer-acting effects
Predictor Treatment Average Per Duration Of Action	Treatment cumulative over duration of action	Cumulative treatment dose

Example: For a user taking Magnesium supplements:

- `predictor_baseline_average_per_day` = 50mg (days not supplementing, dietary only)
- `predictor_treatment_average_per_day` = 400mg (days actively supplementing)
- This reveals the effective treatment dose vs. background exposure

9.5 Relationship Quality Filters

Not all statistically significant relationships are useful. We apply **quality filters** to prioritize actionable findings:

9.5.1 Filter Flags

Flag	Description	Impact on Ranking
Predictor Is Controllable	User can directly modify this predictor (e.g., food, supplements)	Required for actionable recommendations
Outcome Is A Goal	Outcome is something users want to optimize (e.g., mood, energy)	Required for relevance
Plausibly Causal	Plausible biological mechanism exists	Increases confidence
Obvious	Relationship is already well-known (e.g., caffeine → alertness)	May deprioritize for discovery
Boring	Relationship unlikely to interest users	Filters from default views
Interesting Variable Category Pair	Category combination is typically meaningful (e.g., Treatment → Symptom)	Prioritizes for analysis

9.5.2 Boring Relationship Definition

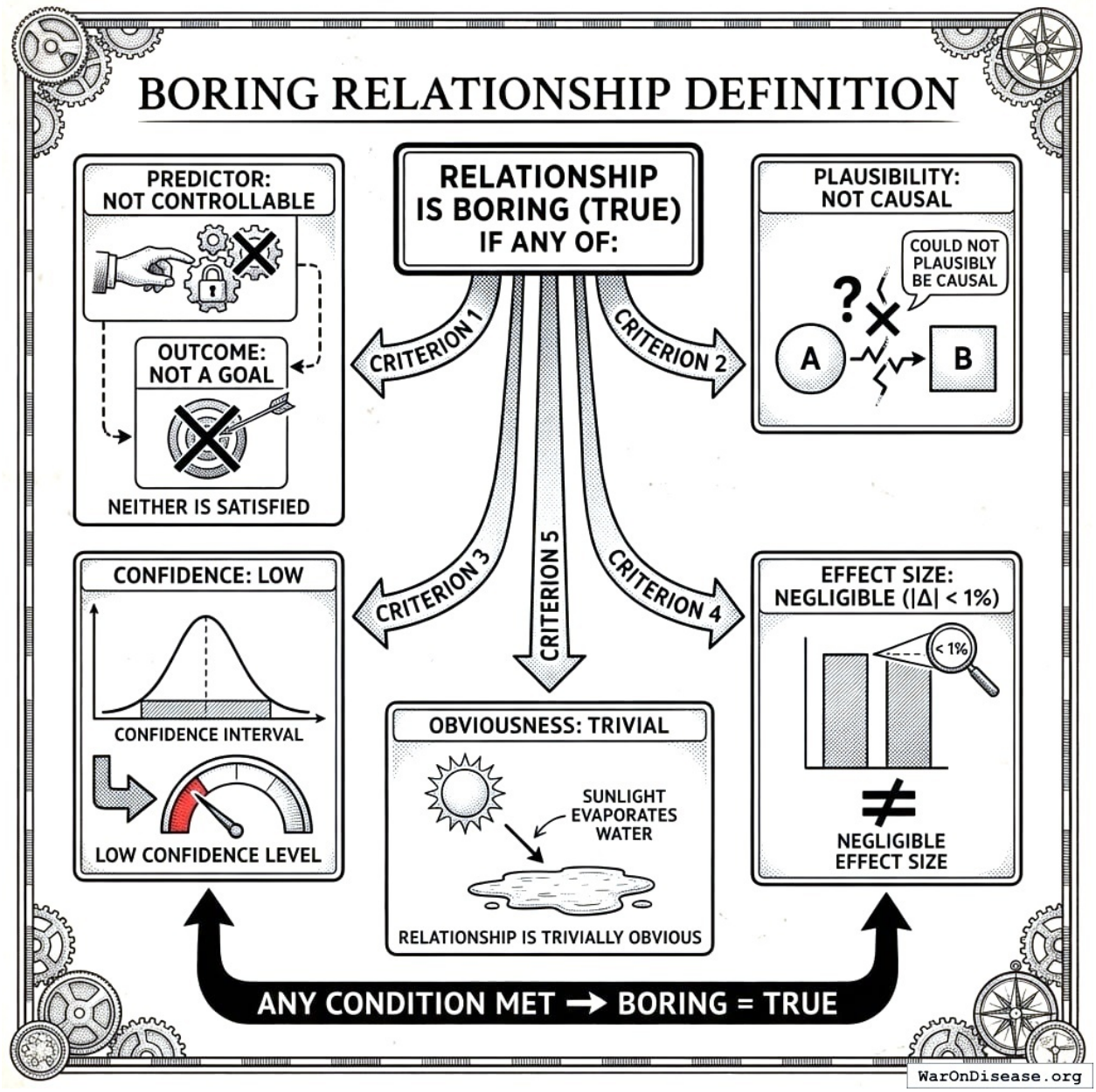

Figure 24: Five ways to tell if your data is too boring to bother with. Science has a spam filter now.

A relationship is marked `boring = TRUE` if ANY of:

- Predictor is not controllable AND outcome is not a goal
- Relationship could not plausibly be causal
- Confidence level is LOW
- Effect size is negligible ($|\Delta| < 1\%$)
- Relationship is trivially obvious

9.5.3 Usefulness and Causality Voting

Users can vote on individual relationships:

Vote Type	Values	Purpose
Usefulness Vote	-1, 0, 1	Whether knowledge of this relationship is useful
Causality Vote	-1, 0, 1	Whether there's a plausible causal mechanism

Aggregate votes contribute to the PIS plausibility weight (w).

9.6 Variable Valence

Valence indicates whether higher values of a variable are inherently good, bad, or neutral:

Valence	Meaning	Examples
Positive	Higher is better	Energy, Sleep Quality, Productivity
Negative	Lower is better	Pain, Anxiety, Fatigue
Neutral	Direction depends on context	Heart Rate, Weight

9.6.1 Impact on Interpretation

Valence affects how we interpret correlation direction:

Predictor-Outcome Valence	Positive Correlation	Negative Correlation
Positive → Positive	Both improve together	Trade-off
Positive → Negative	Predictor worsens outcome	Predictor improves outcome
Treatment → Negative Symptom	Side effect	Therapeutic effect

Example: A positive correlation between Sertraline and Depression Severity is BAD (depression has negative valence, so lower is better). The same positive correlation between Sertraline and Energy would be GOOD.

9.7 Temporal Parameter Optimization

We optimize `onset_delay` ( ) and `duration_of_action` ( ) to find the temporal parameters that maximize correlation strength:

9.7.1 Stored Optimization Data

Field	Description
Correlations Over Delays	Pearson r values for various onset delays
Correlations Over Durations	Pearson r values for various durations of action

Field	Description
Onset Delay With Strongest Pearson Correlation	Optimal value
Pearson Correlation With No Onset Delay	Baseline r for immediate effect
Average Forward Pearson Correlation Over Onset Delays	Mean r across all tested delays
Average Reverse Pearson Correlation Over Onset Delays	Mean reverse r across delays

9.7.2 Optimization Grid

For each predictor-outcome pair, we test:

- **Onset delays**: 0, 30min, 1hr, 2hr, 4hr, 8hr, 12hr, 24hr, 48hr, 72hr…
- **Durations**: 1hr, 4hr, 12hr, 24hr, 48hr, 72hr, 1 week, 2 weeks…

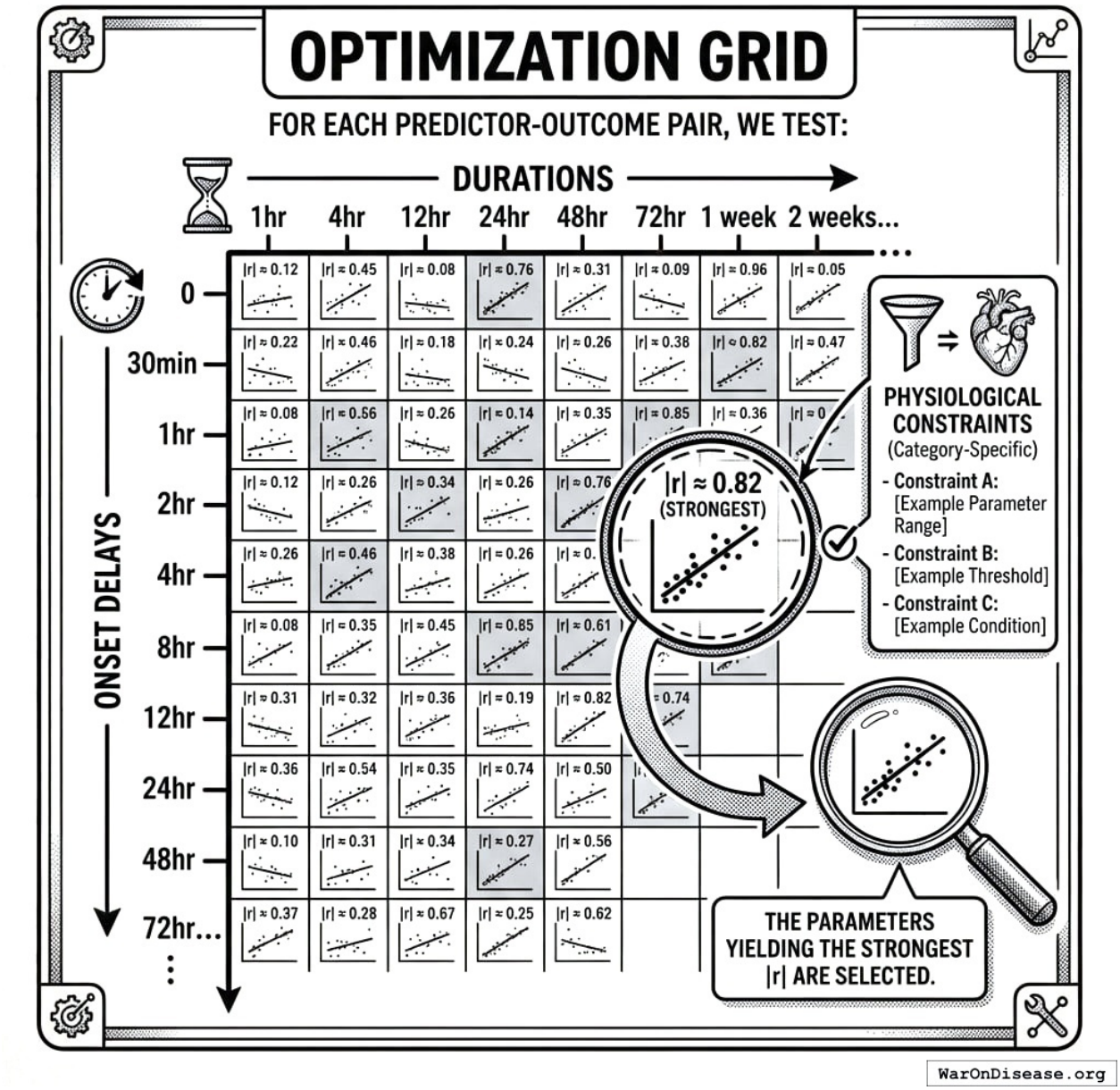

Figure 25: A spreadsheet where every cell represents how long to wait and how long to watch for effects. Somewhere in this grid is the truth. The computer checks every box.

The parameters yielding the strongest |r| are selected, subject to category-specific physiological constraints.

9.7.3 Overfitting Protection

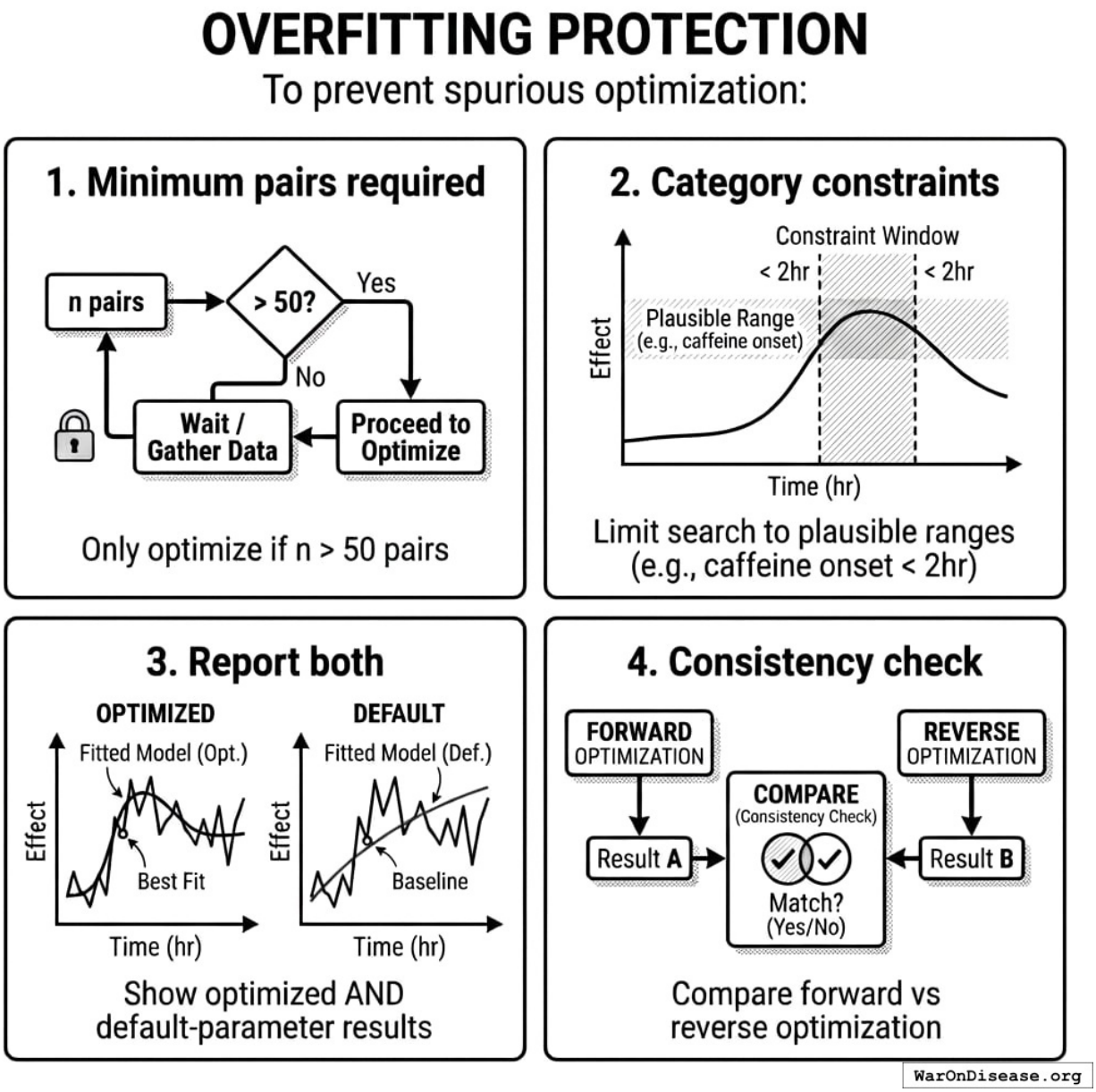

Figure 26: Four ways to stop the computer from seeing patterns that don't exist, like your brain does with clouds.

To prevent spurious optimization: 1. **Minimum pairs required**: Only optimize if n $>$ 50 pairs 2. **Category constraints**: Limit search to plausible ranges (e.g., caffeine onset $<$ 2hr) 3. **Report both**: Show optimized AND default-parameter results 4. **Consistency check**: Compare forward vs reverse optimization

9.8 Spearman Rank Correlation

In addition to Pearson correlation, we compute **Spearman rank correlation** (`forward_spearman_correlation_co`) for robustness:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Where d_i = difference in ranks for each pair.

Advantages over Pearson:

- Robust to outliers
- Captures monotonic (not just linear) relationships
- Less affected by skewed distributions

When to prefer Spearman:

- Outcome has skewed distribution (e.g., symptom severity with many zeros)
- Relationship is monotonic but non-linear (e.g., diminishing returns)
- Data contains outliers from measurement errors

10 Outcome Label Generation

10.1 Predictor Analysis Reports

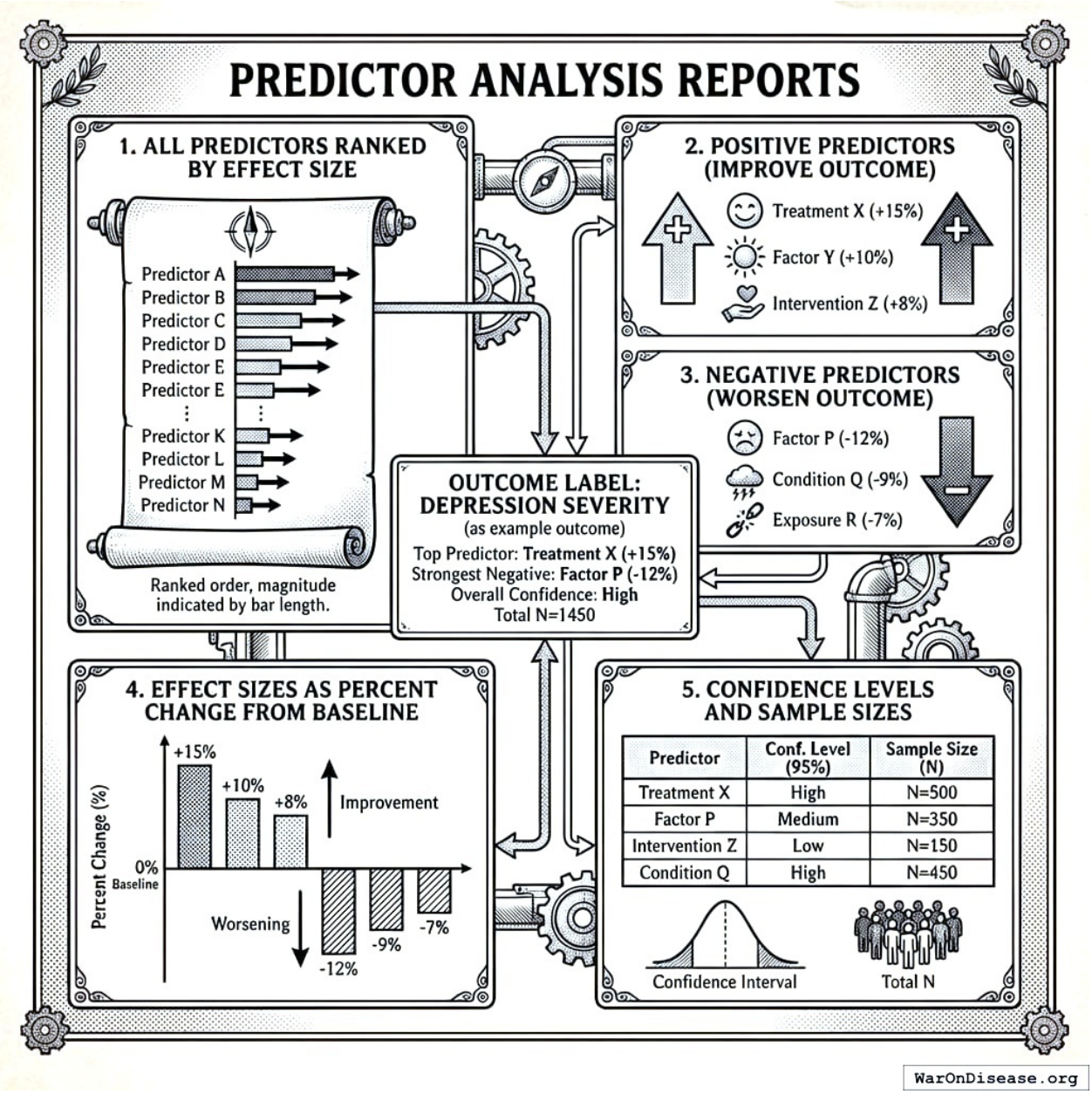

Figure 27: Everything that makes your disease better or worse, ranked from most helpful to most harmful. Like a scoreboard for your organs.

For each outcome variable (e.g., Depression Severity), we generate comprehensive "outcome labels" showing:

1. **All predictors ranked by effect size**
2. **Positive predictors** (treatments/factors that improve the outcome)
3. **Negative predictors** (treatments/factors that worsen the outcome)

4. **Effect sizes as percent change from baseline**
5. **Confidence levels and sample sizes**

10.2 Report Structure

```
Outcome Label: [Outcome Variable Name]
Population: N = [number] participants
Total Studies: [number] treatment-outcome pairs analyzed

POSITIVE EFFECTS (Treatments predicting IMPROVEMENT)
================================================
Rank | Treatment | Effect Size | 95% CI | N | Confidence
-----|-----------|-------------|--------|---|------------
1    | Treatment A | +23.5% | [18.2, 28.8] | 1,247 | High
2    | Treatment B | +18.2% | [12.1, 24.3] | 892 | High
3    | Treatment C | +12.7% | [8.3, 17.1] | 2,103 | High
...

NEGATIVE EFFECTS (Treatments predicting WORSENING)
================================================
Rank | Treatment | Effect Size | 95% CI | N | Confidence
-----|-----------|-------------|--------|---|------------
1    | Treatment X | -15.3% | [-20.1, -10.5] | 567 | Medium
2    | Treatment Y | -8.7% | [-12.3, -5.1] | 1,892 | High
...

NO SIGNIFICANT EFFECT
====================
[List of treatments with |Δ| < threshold or p > 0.05]
```

10.3 Category-Specific Analysis

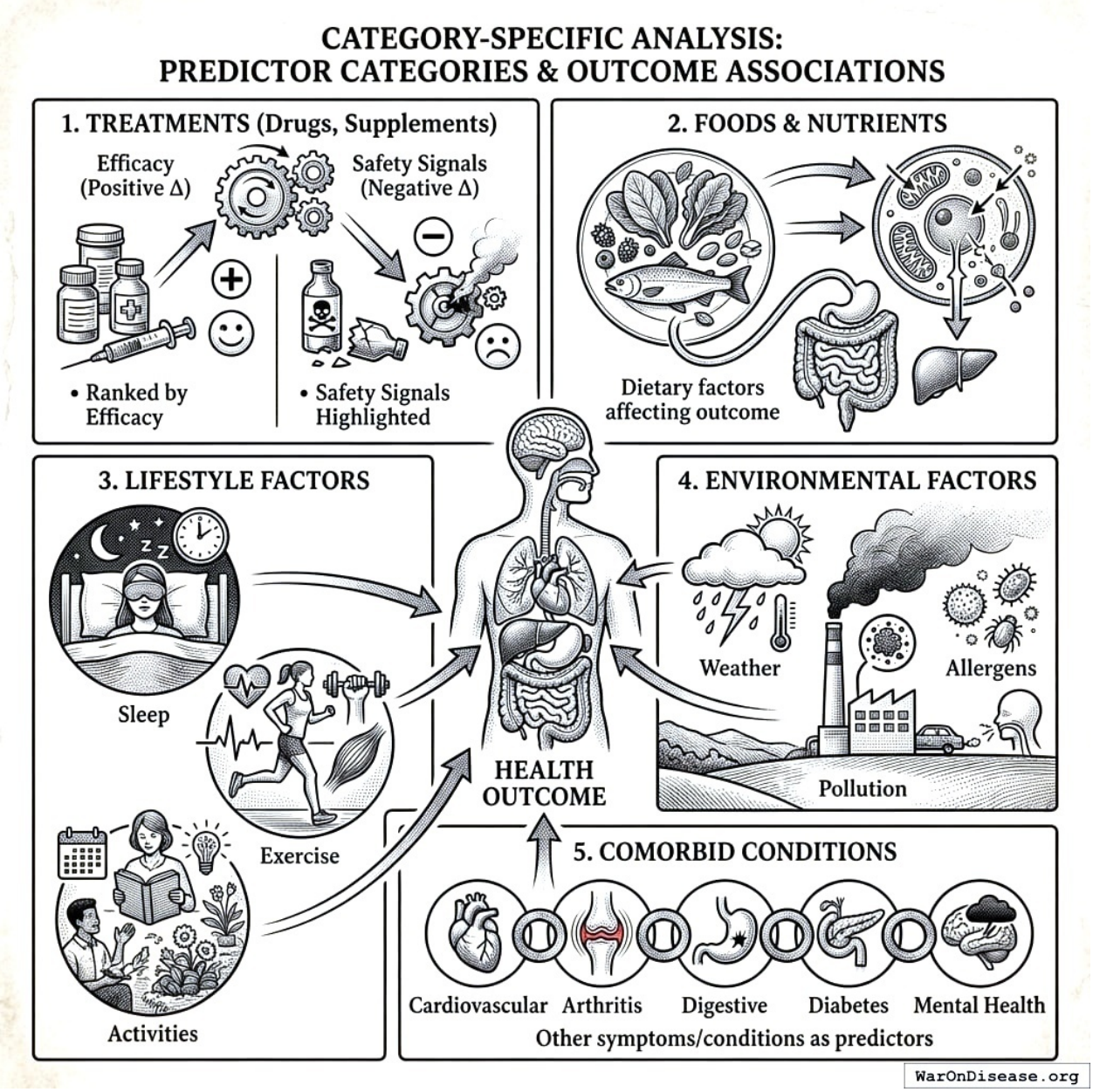

Figure 28: Five categories of things that affect your health: pills, food, habits, air, and other diseases you already have. Medicine filed everything into folders.

Reports are organized by predictor category:

1. **Treatments** (Drugs, Supplements)
   - Ranked by efficacy (positive Δ)
   - Safety signals highlighted (negative Δ)
2. **Foods & Nutrients**
   - Dietary factors affecting outcome

3. **Lifestyle Factors**
    - Sleep, exercise, activities
4. **Environmental Factors**
    - Weather, pollution, allergens
5. **Comorbid Conditions**
    - Other symptoms/conditions as predictors

10.4 Verification Status

Each study is classified by verification status:

Status	Icon	Description
Verified		Up-voted by users; data reviewed and valid
Unverified	?	Awaiting review
Flagged		Down-voted; potential data quality issues

10.5 Outcome Labels vs. FDA Drug Labels

Traditional FDA drug labels are **per-drug documents** that list qualitative adverse events and indications based on pre-market trials. They are static (updated infrequently), qualitative ("may cause drowsiness"), and organized around the drug rather than the patient's condition.

Outcome Labels invert this paradigm: they are **per-outcome documents** that rank all treatments by quantitative effect size for a given health outcome. They are dynamic (updated continuously as data arrives), quantitative ("↓24.7% depression severity"), and organized around what the patient wants to optimize. This enables patients and clinicians to answer the question: "What works best for my condition?" This is a question traditional drug labels cannot answer.

10.6 Worked Example: Complete Outcome Label

The following shows a complete outcome label for depression, demonstrating how treatments are ranked by effect size with confidence intervals:

OUTCOME LABEL: Depression Severity

Based on 47,832 participants tracking depression outcomes Last updated: 2026-01-04 | Data period: 2020-2026

Treatments Improving Depression (ranked by effect size Δ)

Table 1: Treatments associated with depression improvement. Negative effect indicates symptom reduction.

Rank	Treatment	Effect	95% CI	N	PIS	Optimal Dose
1	Exercise	−31.2%	[27.1, 35.3]	12,847	0.67	45 min/day
2	Bupropion	−28.3%	[22.1, 34.5]	2,847	0.54	300mg
3	Sertraline	−24.7%	[19.8, 29.6]	5,123	0.51	100mg
4	Sleep (7-9 hrs)	−22.1%	[18.4, 25.8]	31,204	0.48	8.2 hrs
5	Venlafaxine	−21.2%	[15.3, 27.1]	1,892	0.44	150mg

Rank	Treatment	Effect	95% CI	N	PIS	Optimal Dose
6	Omega-3	−18.9%	[14.2, 23.6]	4,521	0.38	2000mg EPA+DHA
7	Meditation	−16.4%	[12.1, 20.7]	8,932	0.35	20 min/day
8	Fluoxetine	−15.8%	[11.2, 20.4]	3,456	0.33	40mg
9	Vitamin D	−12.3%	[8.7, 15.9]	6,789	0.28	4000 IU
10	Social interaction	−11.7%	[8.2, 15.2]	9,234	0.26	3+ hrs/day

Treatments Worsening Depression (safety signals)

Table 2: Treatments associated with depression worsening. Positive effect indicates symptom increase.

Rank	Treatment	Effect	95% CI	N	PIS	Note
1	Alcohol (>2/day)	+23.4%	[18.9, 27.9]	7,234	0.52	Dose-dependent
2	Sleep deprivation	+19.8%	[15.2, 24.4]	14,521	0.47	<6 hrs/night
3	Social isolation	+15.2%	[11.3, 19.1]	5,892	0.38	<1 hr/day
4	Refined sugar	+8.7%	[5.2, 12.2]	11,234	0.24	>50g/day

No Significant Effect ($|\Delta| < 5\%$ or $p > 0.05$): Multivitamin, Probiotics, B-complex, Magnesium (for depression specifically), Ashwagandha, 5-HTP, SAMe, St. John's Wort[1]

Legend: PIS = Predictor Impact Score (0-1 scale, higher = stronger evidence); Optimal Dose = V_{high} for positive valence outcomes and V_{low} for negative valence outcomes

Interpretation: This outcome label shows that for depression, exercise and sleep optimization rival or exceed pharmaceutical interventions in effect size, with stronger evidence bases (higher N). Bupropion and Sertraline lead among medications. The safety signals section highlights modifiable risk factors that worsen depression.

[1]St. John's Wort shows high heterogeneity (some responders, some non-responders)

Figure 29: Outcome Labels show quantitative effect sizes, sample sizes, and confidence intervals for each treatment, like nutrition facts for drugs

11 Treatment Ranking System

11.1 Within-Category Rankings

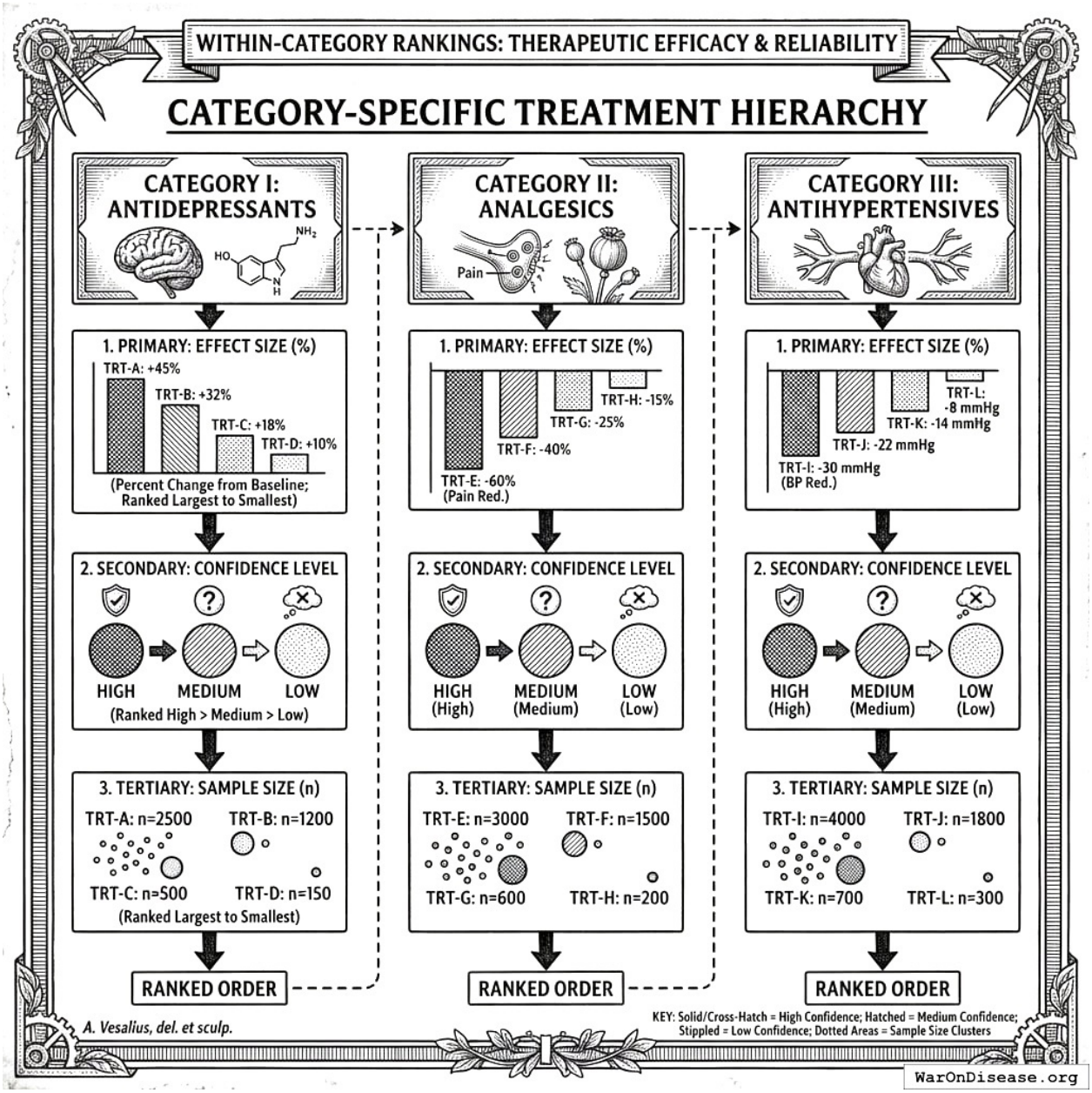

Figure 30: Treatments ranked by: does it work (most important), are we sure (pretty important), and how many people did we watch (least important). Revolutionary prioritization.

For each therapeutic category (e.g., Antidepressants), treatments are ranked by:

1. **Primary**: Effect size (percent change from baseline)
2. **Secondary**: Confidence level (High > Medium > Low)
3. **Tertiary**: Sample size

11.2 Ranking Algorithm

For each treatment T in a therapeutic category, we compute a composite ranking score:

$$\text{RankScore}_T = \bar{\Delta}_T \times w_{\text{confidence}} \times \text{PIS}_T$$

where $\bar{\Delta}_T$ is the mean effect size across participants, $w_{\text{confidence}}$ is the confidence weight (see Table 3), and PIS_T is the Predictor Impact Score. Treatments are sorted by descending rank score.

11.3 Confidence Weighting

Table 3: Confidence weighting schema for treatment ranking.

Confidence Level	Weight (w)	Criteria
High	1.0	$p < 0.01$ OR $N > 100$ OR pairs > 500
Medium	0.7	$p < 0.05$ OR $N > 10$ OR pairs > 100
Low	0.4	Meets minimum thresholds only

11.4 Comparative Effectiveness Display

Table 4 illustrates how treatments within a therapeutic category are presented to users.

Table 4: Antidepressants ranked by efficacy for depression. Negative effect indicates symptom reduction.

Rank	Treatment	Effect (Δ)	95% CI	N	Confidence
1	Bupropion 300mg	-28.3%	[22.1, 34.5]	2,847	High
2	Sertraline 100mg	-24.7%	[19.8, 29.6]	5,123	High
3	Venlafaxine 150mg	-21.2%	[15.3, 27.1]	1,892	High
4	Fluoxetine 40mg	-18.9%	[13.2, 24.6]	3,456	High

12 Safety and Efficacy Quantification

12.1 Safety Signal Detection

Adverse Effect Identification: Safety signals are identified through (1) negative correlations between treatment and beneficial outcomes, and (2) positive correlations between treatment and harmful outcomes.

Table 5: Example safety signal report showing potential adverse effects with statistically significant positive correlations to harmful outcomes.

Outcome	Effect (Δ)	95% CI	Plausibility	Action
Fatigue	$+18.3\%$	[12.1, 24.5]	High (known sedation)	Monitor
Nausea	$+15.7\%$	[8.9, 22.5]	High (GI effects)	Monitor
Weight Gain	$+8.2\%$	[4.1, 12.3]	Medium	Long-term monitoring

Outcome	Effect (Δ)	95% CI	Plausibility	Action
Anxiety	+6.5%	[2.1, 10.9]	Low (paradoxical)	Investigate

12.2 Efficacy Signal Detection

Therapeutic Effect Identification: Efficacy signals are identified through (1) positive correlations between treatment and beneficial outcomes, and (2) negative correlations between treatment and harmful outcomes (symptom reduction).

Table 6: Example efficacy signal report showing therapeutic effects with statistically significant correlations.

Outcome	Effect (Δ)	95% CI	Indication	Evidence
Depression	−24.7%	[19.8, 29.6]	Primary	Strong
Anxiety	−18.2%	[12.3, 24.1]	Secondary	Strong
Sleep Quality	+15.3%	[10.1, 20.5]	Secondary	Moderate
Energy	+12.1%	[7.2, 17.0]	Secondary	Moderate

12.3 Benefit-Risk Assessment

Net Clinical Benefit Score:

$$\text{NCB} = \sum_{i \in \text{benefits}} w_i \cdot |\Delta_i| - \sum_{j \in \text{risks}} w_j \cdot |\Delta_j|$$

where w represents importance weights assigned by clinical relevance.

Example: Sertraline 100mg Benefit-Risk Profile

Table 7: Benefit-risk components for Sertraline 100mg.

Benefits	Effect	Weight	Risks	Effect	Weight
Depression	−24.7%	1.0	Nausea	+8.3%	0.3
Anxiety	−18.2%	0.8	Insomnia	+5.1%	0.4
			Sexual dysfunction	+12.7%	0.5

Weighted Summary: Benefits = 39.26, Risks = 8.93, **Net Clinical Benefit = +30.33** (favorable profile for depression/anxiety)

13 Addressing the Bradford Hill Criteria

The Bradford Hill criteria[146] provide the foundational framework for assessing causation from observational data. This section details how our framework addresses each criterion.

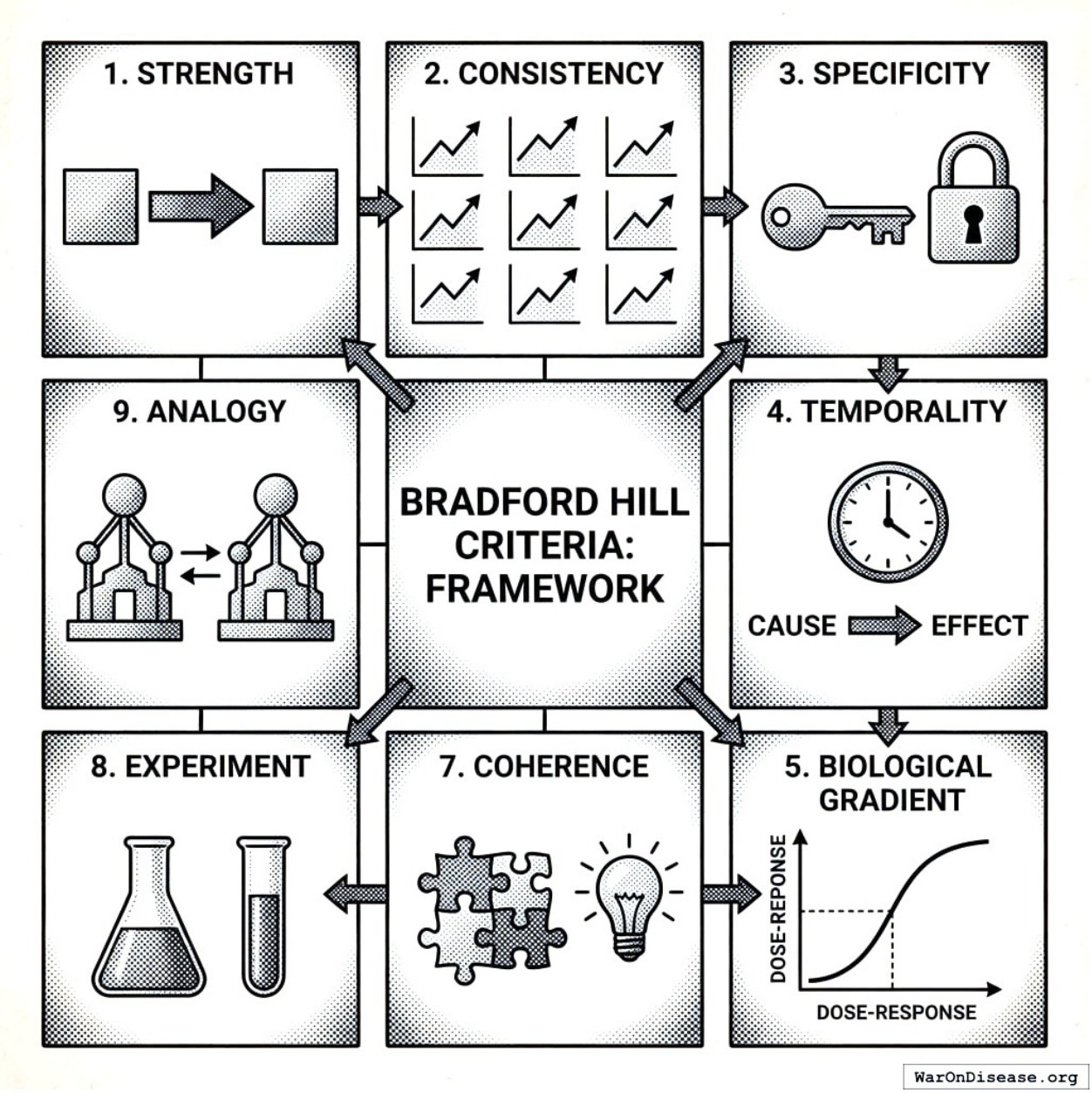

Figure 31: Nine ways to tell if A causes B or if you're just making things up. Bradford Hill wrote them down in 1965. You've been ignoring them.

13.1 Complete Criteria Mapping

Criterion	How Addressed	Quantitative Metric	In PIS?
Strength	Effect size magnitude	Pearson r, $\Delta\%$	Yes
Consistency	Cross-participant aggregation	N, n, SE, CI	Yes
Specificity	Category appropriateness	Interest factor	Yes

Criterion	How Addressed	Quantitative Metric	In PIS?
Temporality	Onset delay requirement	$\delta > 0$ enforced	Yes
Biological Gradient	Dose-response analysis	Gradient coefficient	Yes
Plausibility	Community voting	Up/down votes	Yes
Coherence	Literature cross-reference	Narrative	No
Experiment	N-of-1 natural experiments	Study design	No
Analogy	Similar variable comparison	Narrative	No

13.2 Quantitative Criteria Details

Strength:

- Reports Pearson r with classification (very strong: 0.8, strong: 0.6, moderate: 0.4, weak: 0.2, very weak: <0.2)
- Example: "There is a moderately positive (R = 0.45) relationship between Sertraline and Depression improvement."

Consistency:

- Reports N participants, n paired measurements
- Notes that spurious associations naturally dissipate as participants modify behaviors based on non-replicating findings

Temporality:

- Onset delay δ explicitly encodes treatment-to-effect lag
- Forward vs. reverse correlation comparison identifies potential reverse causality

Plausibility:

- Users vote on biological mechanism plausibility
- Weighted average contributes to ranking
- Crowd-sources expert and patient knowledge

14 Validation and Quality Assurance

14.1 User Voting System

Each study can receive user votes:

Vote	Meaning	Effect
Up-vote (+1)	Data appears valid, relationship plausible	Included in verified results
Down-vote (-1)	Data issues or implausible relationship	Flagged for review

Vote	Meaning	Effect
No vote	Not yet reviewed	Included in unverified results

14.2 Automated Quality Checks

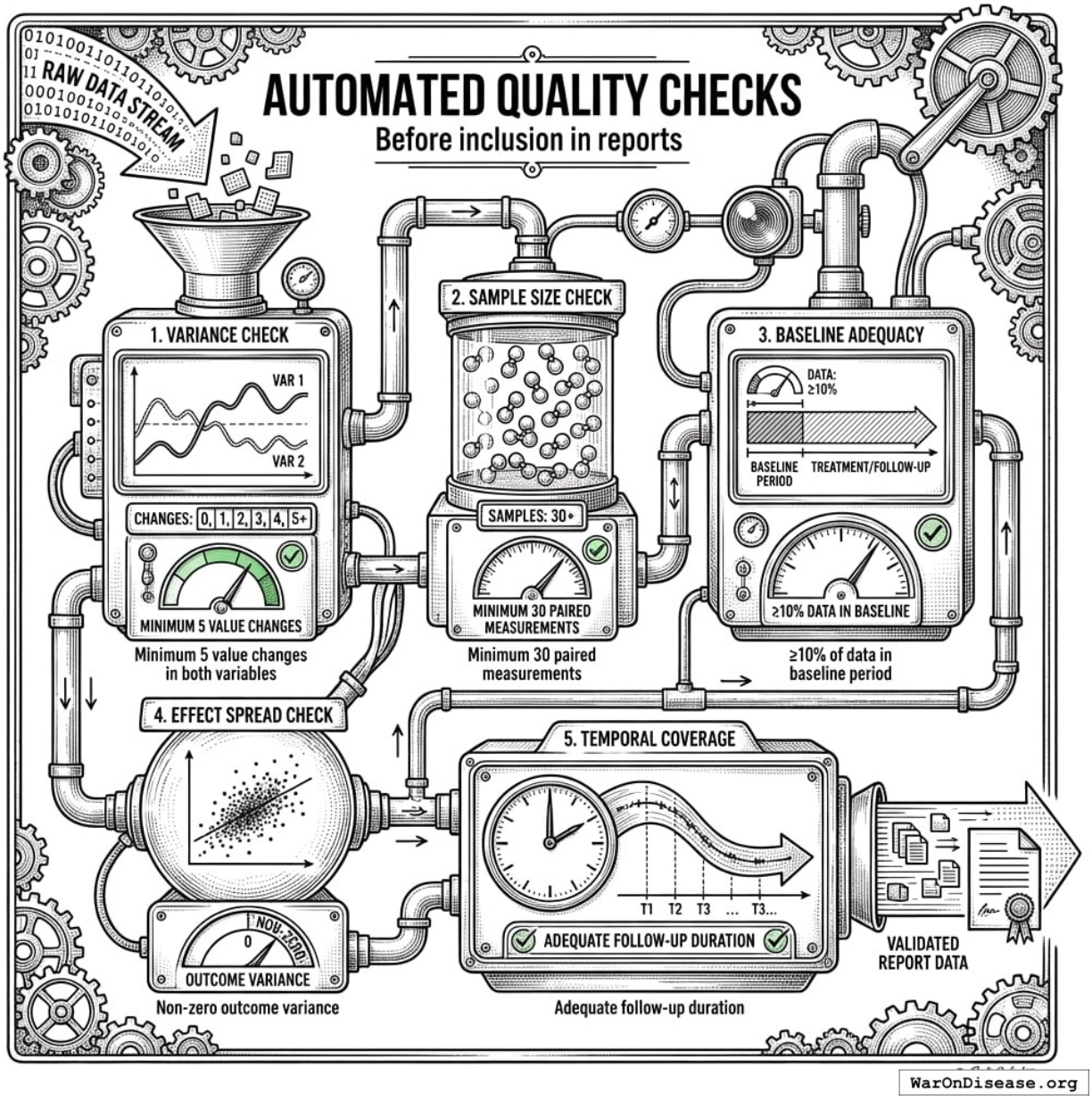

Figure 32: Five checkpoints where bad data gets thrown out. It's airport security, but for numbers.

Before inclusion in reports:

1. **Variance check**: Minimum 5 value changes in both variables
2. **Sample size check**: Minimum 30 paired measurements

3. **Baseline adequacy**: 10% of data in baseline period
4. **Effect spread check**: Non-zero outcome variance
5. **Temporal coverage**: Adequate follow-up duration

14.3 Flagged Study Handling

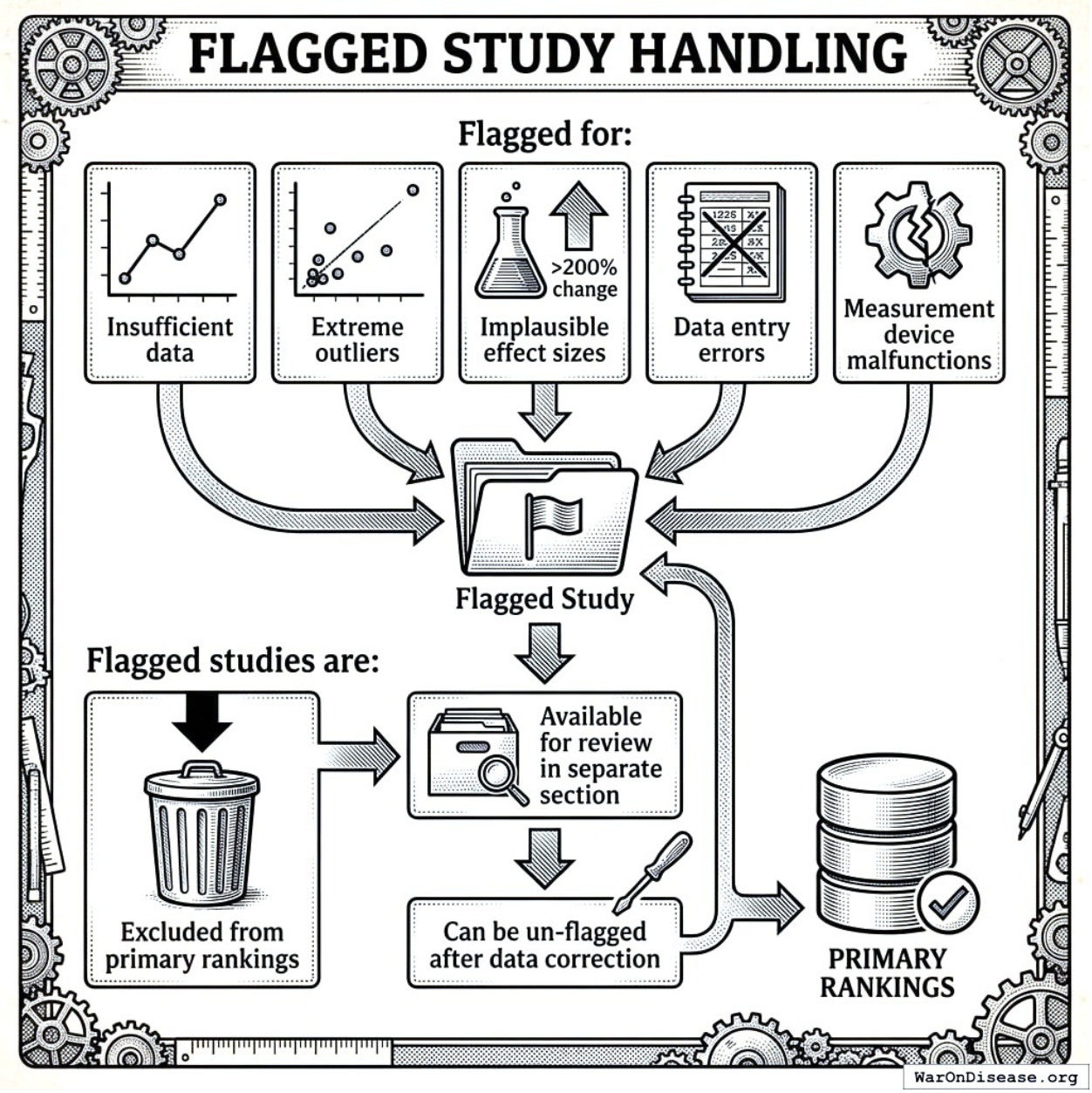

Figure 33: Studies can get kicked out for being terrible, then let back in if they fix their mistakes. Academic probation, but for data.

Studies may be flagged for:

- Insufficient data

- Extreme outliers
- Implausible effect sizes (>200% change)
- Data entry errors
- Measurement device malfunctions

Flagged studies are:

- Excluded from primary rankings
- Available for review in separate section
- Can be un-flagged after data correction

15 Stage 2: Pragmatic Trial Confirmation

The short version: Observational data can find promising signals, but only randomized trials can prove causation. The good news: we don't need expensive, slow traditional trials. A meta-analysis of 108 embedded pragmatic trials[141] shows that "pragmatic" trials (simple randomization embedded in routine care) can validate treatments at 44.1x (95% CI: 39.4x-89.1x) lower cost. We use cheap observational analysis (Stage 1) to filter millions of possibilities down to the top candidates, then confirm the best ones with pragmatic trials (Stage 2). Result: validated treatment recommendations at a fraction of current cost.

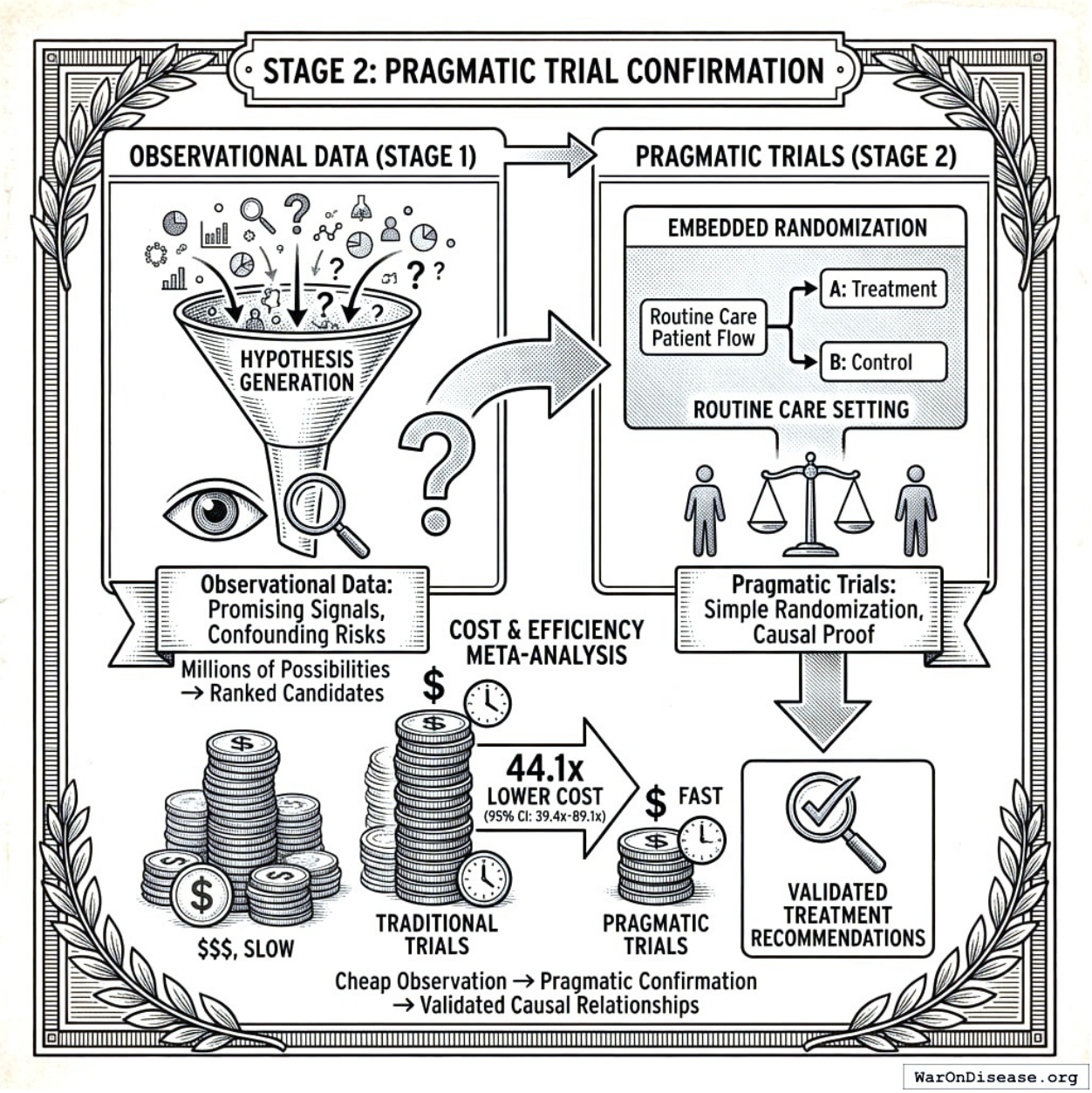

Figure 34: Stage 1: Computers watch everyone and get suspicious about patterns. Stage 2: Humans run cheap experiments to see if the computers were right or hallucinating.

The observational methodology described in Sections 1-11 generates ranked hypotheses about treatment-outcome relationships. While powerful for signal detection and hypothesis generation, observational data alone cannot establish causation due to unmeasured confounding. This section describes how **pragmatic clinical trials** serve as the confirmation layer, transforming promising observational signals into validated causal relationships.

15.1 The Two-Stage Pipeline

Our complete methodology operates as a two-stage pipeline:

Table 8: Two-stage pipeline summary.

Stage	Method	Cost	Purpose	Output
Stage 1: Signal Detection	Aggregated N-of-1 observational analysis	~$0.1 (95% CI: $0.03-$1)/patient	Hypothesis generation	Ranked PIS signals
Stage 2: Causal Confirmation	Pragmatic randomized trials	~$929 (95% CI: $97-$3K)/patient	Causation proof	Validated effect sizes

This design leverages the complementary strengths of each approach:

- **Stage 1** scales to millions of treatment-outcome pairs at minimal cost, identifying the most promising candidates
- **Stage 2** applies experimental rigor to confirm causation for high-priority signals

15.2 Pragmatic Trial Methodology

Pragmatic trials differ fundamentally from traditional Phase III trials. A Harvard meta-analysis of 108 embedded pragmatic trials found median costs of only $97/patient, with even conservative implementations like ADAPTABLE achieving $929 (95% CI: $929-$1.4K)/patient[1,141]:

Table 9: Pragmatic vs. traditional Phase III trials.

Dimension	Traditional Phase III	Pragmatic Trial (Evidence-Based)
Cost per patient	$41K (95% CI: $20K-$120K)	$929 (95% CI: $97-$3K) (median $97-929)[2]
Time to results	3-7 years	3-6 months
Patient population	Homogeneous (strict exclusion)	Real-world (minimal exclusion)
Setting	Specialized research centers	Routine clinical care
Data collection	Extensive case report forms	Minimal essential outcomes
Randomization	Complex stratification	Simple 1:1 or 1:1:1
Sample size	Hundreds to thousands	Thousands to tens of thousands

Multiple large-scale pragmatic trials have demonstrated this model's effectiveness. The Oxford RECOVERY trial enrolled 49,000 patients across 186 hospitals, evaluating 12 treatments and finding a life-saving result (dexamethasone) in 100 days, saving 1 million lives (95% CI: 500 thousand lives-2 million lives) globally[89]. The PCORnet ADAPTABLE trial enrolled 15,076 patients across 40 clinical sites at $929 (95% CI: $929-$1.4K)/patient[1]. These are not isolated successes. The Harvard meta-analysis shows this efficiency is reproducible across therapeutic areas[141].

[2]Meta-analysis of 108 trials found $97 median; ADAPTABLE trial achieved $929; RECOVERY achieved $500. We use the conservative ADAPTABLE estimate.

15.3 Signal-to-Trial Prioritization

Not all observational signals warrant pragmatic trial confirmation. We propose a **Trial Priority Score (TPS)** combining:

$$TPS = PIS \times \sqrt{\text{DALYs}_{\text{addressable}}} \times \text{Novelty} \times \text{Feasibility}$$

Where:

- **PIS**: Predictor Impact Score from Stage 1 (higher = stronger signal)
- $\text{DALYs}_{\textbf{addressable}}$: Disease burden addressable by the treatment
- **Novelty**: Inverse of existing evidence (new signals prioritized)
- **Feasibility**: Practical considerations (drug availability, safety profile, cost)

Signals in the top 0.1-1% by TPS are candidates for pragmatic trial confirmation.

15.4 Comparative Effectiveness Randomization

For treatments already in clinical use, we employ **comparative effectiveness** designs following the ADAPTABLE trial model[1]:

1. **Embedded randomization**: Randomization occurs within routine care visits
2. **Minimal disruption**: Patients receive standard care with random assignment between active comparators
3. **Real-world endpoints**: Primary outcomes are events captured in EHR (mortality, hospitalization, symptom resolution)
4. **Large simple design**: Thousands of patients, minimal per-patient data collection

Example protocol for a high-PIS signal (Treatment A vs. Treatment B for Outcome X):

Table 10: Example pragmatic trial protocol for comparative effectiveness.

Parameter	Specification
Eligibility	Patients with Condition Y initiating treatment for Outcome X
Randomization	1:1 to Treatment A vs. Treatment B
Primary endpoint	Change in Outcome X at 90 days
Data collection	Baseline characteristics (EHR), outcome at 90 days (patient-reported or EHR)
Sample size	2,000 patients (1,000 per arm)
Cost	~$1.9M total ($929 (95% CI: $97-$3K)/patient)
Timeline	6-12 months

15.5 Feedback Loop: Trial Results Improve Observational Models

Pragmatic trial results feed back to improve Stage 1 methodology:

1. **Calibration**: Compare observational effect sizes to randomized effect sizes; develop correction factors

2. **Confounding identification**: Trials where observational and randomized effects diverge identify confounders
3. **Subgroup discovery**: Trial heterogeneity analysis identifies responder populations, improving PIS stratification
4. **Hyperparameter validation**: Optimal onset delays and durations validated against experimental ground truth

This creates a **learning health system** where observational and experimental evidence continuously refine each other.

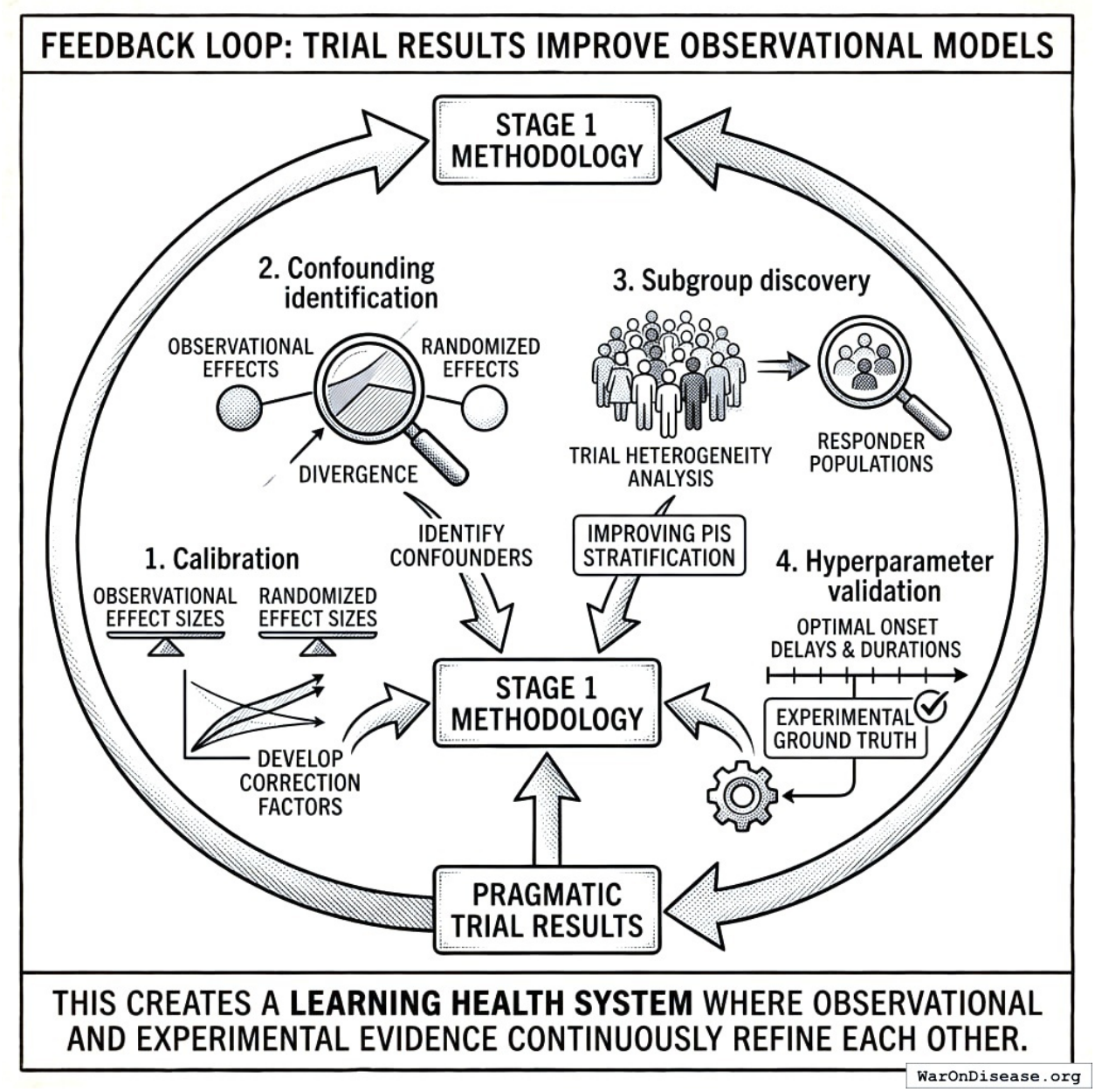

Figure 35: A loop where real life teaches experiments what to test, and experiments teach real life what works. It's a circle, which means it never stops, which terrifies administrators.

15.6 Output: Validated Outcome Labels

The two-stage pipeline produces **validated outcome labels** combining observational and experimental evidence. Table 11 shows the data elements captured for each treatment-outcome pair.

Table 11: Validated outcome label data structure.

Component	Field	Description	Example
Identification	Treatment	Intervention name and dose	Vitamin D 2000 IU
	Outcome	Health outcome measured	Depression Severity
Stage 1 (Observational)	Δ_{obs}	Observational effect size	-12%
	CI_{obs}	95% confidence interval	$[-15\%, -9\%]$
	N_{obs}	Number of participants	45,000
	PIS	Predictor Impact Score	0.72
Stage 2 (Experimental)	Δ_{exp}	Randomized trial effect	-8%
	CI_{exp}	Trial confidence interval	$[-12\%, -4\%]$
	N_{exp}	Trial participants	3,000
	Trial ID	Registry identifier	DFDA-VIT-D-001
Combined	Evidence Grade	Validation status	Validated/Promising/Signal
	Causal Confidence	Probability of true effect	0-1 scale

Evidence grades:

- **Validated**: Confirmed by pragmatic RCT ($p < 0.05$, consistent direction)
- **Promising**: High PIS (>0.6), awaiting or in trial
- **Signal**: Moderate PIS (0.3-0.6), hypothesis only

16 Limitations and How They're Addressed

The two-stage design addresses the fundamental limitations of purely observational pharmacovigilance while acknowledging residual constraints.

16.1 Fundamental Limitations: Observational Stage

These limitations apply to Stage 1 (observational analysis) but are addressed by Stage 2 (pragmatic trials):

Table 12: Fundamental observational limitations and trial resolution.

Limitation	Stage 1 Status	Stage 2 Resolution
Cannot prove causation	Hypothesis only	Randomization establishes causation

Limitation	Stage 1 Status	Stage 2 Resolution
Cannot replace RCTs	Generates candidates	Pragmatic trials ARE simplified RCTs
Cannot handle strong confounding	Confounding by indication	Randomization eliminates confounding
Cannot generalize beyond population	Self-selected trackers	Pragmatic trials use real-world populations

16.2 Methodological Weaknesses: Addressed by Two-Stage Design

Table 13: Methodological weaknesses addressed by the two-stage design.

Weakness	Stage 1 Impact	Two-Stage Resolution
Arbitrary baseline definition	Acceptable for signal ranking	Trial uses randomized comparison, no baseline needed
Hyperparameter overfitting	May inflate some correlations	Trial confirms true effect, calibrates models
Self-selection bias	Non-representative sample	Pragmatic trials embed in routine care
Measurement error	Self-report limitations	Trials can use objective endpoints
Hawthorne effect	Tracking changes behavior	Trials embedded in normal care minimize this
Multiple testing	Millions of comparisons	Only top signals proceed to trial (TPS filter)
Temporal confounding	Seasonal/life event effects	Randomization eliminates systematic bias
Confounding by indication	Sicker patients take more treatment	Randomization balances severity

16.3 Residual Limitations

Even with the two-stage design, certain limitations remain:

1. **Resource constraints**: Cannot trial all promising signals; prioritization required
2. **External validity**: Pragmatic trial populations still may not represent all subgroups
3. **Rare outcomes**: Very rare adverse events may require larger observational signals
4. **Behavioral interventions**: Some treatments (diet, exercise) difficult to blind
5. **Long-term effects**: Pragmatic trials typically 6-12 months; decades-long effects require observational follow-up
6. **Interaction effects**: Two-way drug interactions testable; higher-order interactions remain observational
7. **Multiple testing burden**: Stage 1 analyzes millions of treatment-outcome pairs without formal multiple testing correction (e.g., Benjamini-Hochberg FDR control). The TPS filter reduces false positives proceeding to Stage 2, but users should expect a high proportion of Stage 1 signals to be false discoveries. This is by design (cheap observational analysis tolerates false positives because Stage 2 trials filter them out), but consumers of Stage 1 rankings alone should interpret with appropriate skepticism

8. **Self-selection bias**: Participants who track health data differ systematically from the general population (likely healthier, more educated, more health-conscious). Effect sizes may not generalize to non-trackers

16.4 What This Framework CAN Now Do

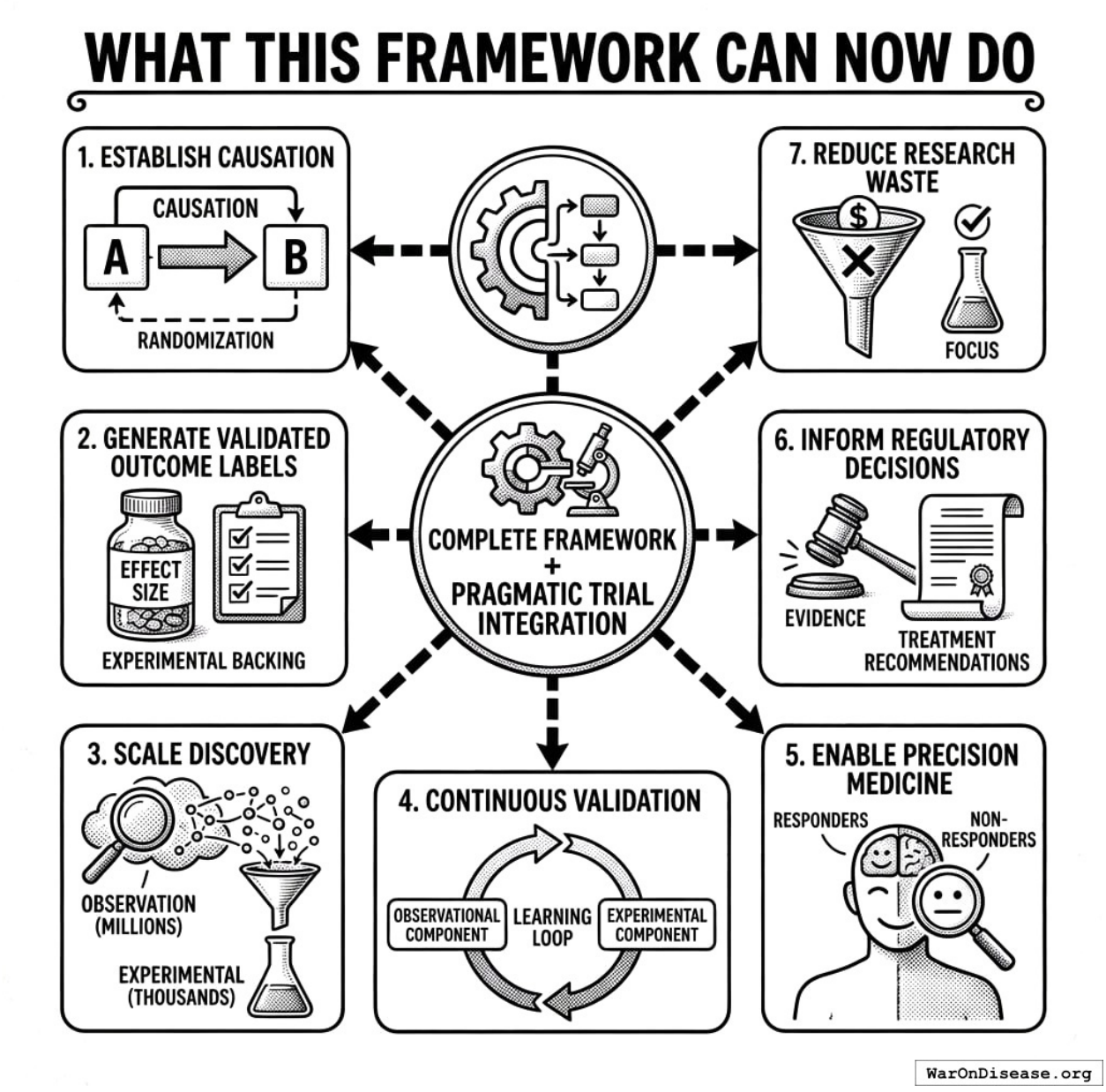

Figure 36: Watch people. Test hunches. Learn things. Watch more people. Test new hunches. Never stop. This is what learning looks like when you automate it.

With pragmatic trial integration, the complete framework can:

1. **Establish causation**: For high-priority signals, randomization proves causal relationships
2. **Generate validated outcome labels**: Quantitative effect sizes with experimental backing

3. **Scale discovery**: Analyze millions of pairs observationally, confirm thousands experimentally
4. **Continuous validation**: Learning loop improves both observational and experimental components
5. **Enable precision medicine**: Subgroup analyses identify responders vs. non-responders
6. **Inform regulatory decisions**: Validated labels provide evidence for treatment recommendations
7. **Reduce research waste**: Focus expensive trials on signals most likely to confirm

17 Implementation Guide

17.1 System Architecture

The processing protocol defines six sequential steps:

1. **Data Ingestion Protocol**: Collects measurements from wearables, health apps, EHR/FHIR systems, manual entry, and environmental sensors
2. **Measurement Normalization**: Standardizes units, timestamps, deduplicates records, and attributes data provenance
3. **Variable Ontology**: Assigns semantic categories, default temporal parameters (δ, τ), and filling value logic
4. **Relationship Analysis Engine**: Generates predictor-outcome pairs, performs temporal alignment, computes correlations, and optimizes hyperparameters
5. **Population Aggregation**: Combines individual N-of-1 analyses, computes confidence intervals, detects heterogeneity and subgroups
6. **Report Generation**: Produces outcome labels, treatment rankings, and safety/efficacy signals

Complete implementation details, database schemas, and reference code are available in the supplementary materials repository.

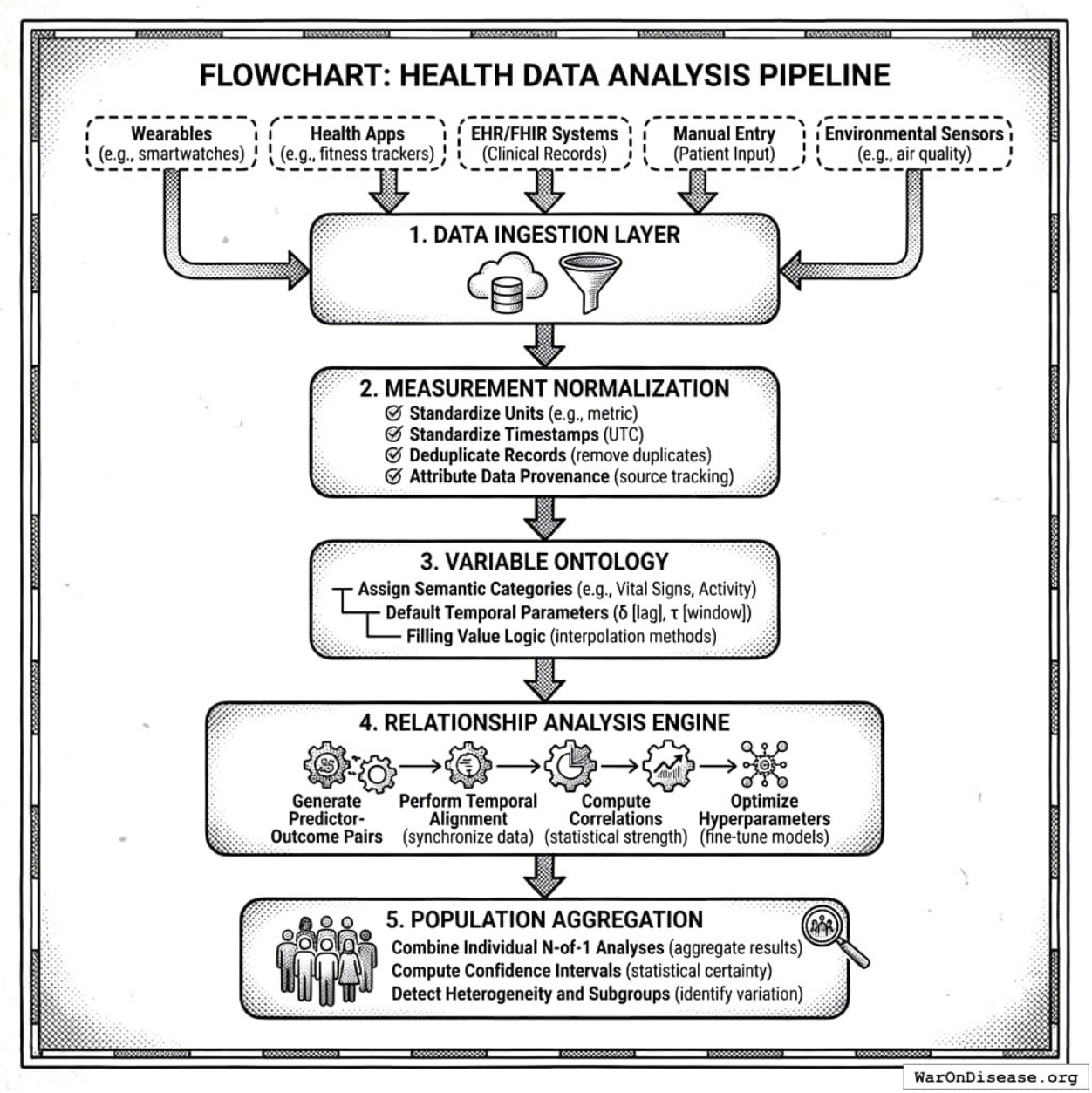

Figure 37: How to turn a billion people's random health facts into useful medical knowledge: a pipeline with more steps than your morning routine.

17.2 Core Algorithm: Pair Generation

Algorithm 1 (Temporal Pair Generation): Given predictor measurements $P = \{(t_j^P, p_j)\}$ and outcome measurements $O = \{(t_k^O, o_k)\}$, onset delay δ, duration of action τ, and optional filling value f:

Case 1 (Predictor has filling value): For each outcome measurement (t_k^O, o_k):

$$p_k = \begin{cases} \frac{1}{|W_k|} \sum_{j \in W_k} p_j & \text{if } W_k \neq \emptyset \\ f & \text{otherwise} \end{cases}$$

where $W_k = \{j : t_k^O - \delta - \tau < t_j^P \leq t_k^O - \delta\}$. Output pair (p_k, o_k).

Case 2 (No filling value): For each predictor measurement (t_j^P, p_j):

$$o_j = \frac{1}{|W_j|} \sum_{k \in W_j} o_k$$

where $W_j = \{k : t_j^P + \delta \leq t_k^O < t_j^P + \delta + \tau\}$. Output pair (p_j, o_j) only if $W_j \neq \emptyset$.

17.3 Core Algorithm: Baseline Separation

Algorithm 2 (Baseline/Follow-up Partition): Given aligned pairs $\{(p_i, o_i)\}_{i=1}^n$:

1. Compute predictor mean: $\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i$
2. Partition into baseline $B = \{(p_i, o_i) : p_i < \bar{p}\}$ and follow-up $F = \{(p_i, o_i) : p_i \geq \bar{p}\}$
3. Compute outcome means: $\mu_B = \mathbb{E}[o \mid (p, o) \in B]$ and $\mu_F = \mathbb{E}[o \mid (p, o) \in F]$
4. Return percent change: $\Delta = \frac{\mu_F - \mu_B}{\mu_B} \times 100$

17.4 Algorithm 3: Predictor Impact Score Calculation

Algorithm 3 (User-Level PIS): Given correlation coefficients, statistical significance, z-score, interest factor, and aggregate PIS:

$$\text{PIS}_{\text{user}} = |r| \times S \times \phi_z \times \phi_{\text{temporal}} \times f_{\text{interest}} + \text{PIS}_{\text{agg}}$$

Procedure:

1. Set $Z_{\text{ref}} = 2$ (reference z-score threshold)
2. Compute strength: $r = |r_{\text{forward}}|$
3. Compute effect magnitude factor: $\phi_z = \frac{|z|}{|z| + Z_{\text{ref}}}$ (or 0.5 if z undefined)
4. Compute temporality factor: $\phi_{\text{temporal}} = \frac{|r_{\text{forward}}|}{|r_{\text{forward}}| + |r_{\text{reverse}}|}$ (or 0.5 if both zero)
5. Return: $r \times S \times \phi_z \times \phi_{\text{temporal}} \times f_{\text{interest}} + \text{PIS}_{\text{agg}}$

Algorithm 4 (Aggregate PIS): Given forward correlation, number of users N, number of pairs n, outcome changes, plausibility votes, and gradient coefficient:

$$\text{PIS}_{\text{agg}} = |r_{\text{forward}}| \times w \times \phi_{\text{users}} \times \phi_{\text{pairs}} \times \phi_{\text{change}} \times \phi_{\text{gradient}}$$

Procedure:

1. Set saturation constants: $N_{\text{sig}} = 10$, $n_{\text{sig}} = 100$, $\Delta_{\text{sig}} = 10\%$
2. Compute user saturation: $\phi_{\text{users}} = 1 - e^{-N/N_{\text{sig}}}$
3. Compute pair saturation: $\phi_{\text{pairs}} = 1 - e^{-n/n_{\text{sig}}}$
4. Compute change spread: $\Delta_{\text{spread}} = |\Delta_{\text{high}} - \Delta_{\text{low}}|$ (minimum 1)
5. Compute change saturation: $\phi_{\text{change}} = 1 - e^{-\Delta_{\text{spread}}/\Delta_{\text{sig}}}$
6. Compute gradient factor: $\phi_{\text{gradient}} = \min(\text{gradient\_coefficient}, 1.0)$
7. Return: $|r_{\text{forward}}| \times w \times \phi_{\text{users}} \times \phi_{\text{pairs}} \times \phi_{\text{change}} \times \phi_{\text{gradient}}$

Algorithm 5 (Interest Factor): Penalize spurious variable pairs:

$$f_{\text{interest}} = f_P \times f_O \times f_{\text{pair}}$$

Procedure:

1. Initialize $f = 1.0$
2. **Predictor penalties**: Divide f by 2 for each: test variable, app/website, address
3. **Outcome penalties**: Divide f by 2 for each: test variable, non-outcome category
4. **Pair penalties**: Divide f by 2 if predictor is non-predictor category; divide by 10 if illogical category pair
5. Return f

17.5 Reference Implementation

A complete reference implementation is available at:

- **Code Repository**: https://github.com/decentralized-fda
- **Documentation**: https://docs.dfda.earth

The repository includes:

- Analysis engine (Python) with full PIS calculation pipeline
- Database schemas (PostgreSQL, SQLite) for variable relationships
- API protocol (OpenAPI 3.0) for data ingestion and querying
- Test vectors matching Appendix D worked example

18 Regulatory Considerations

18.1 Positioning Relative to RCTs

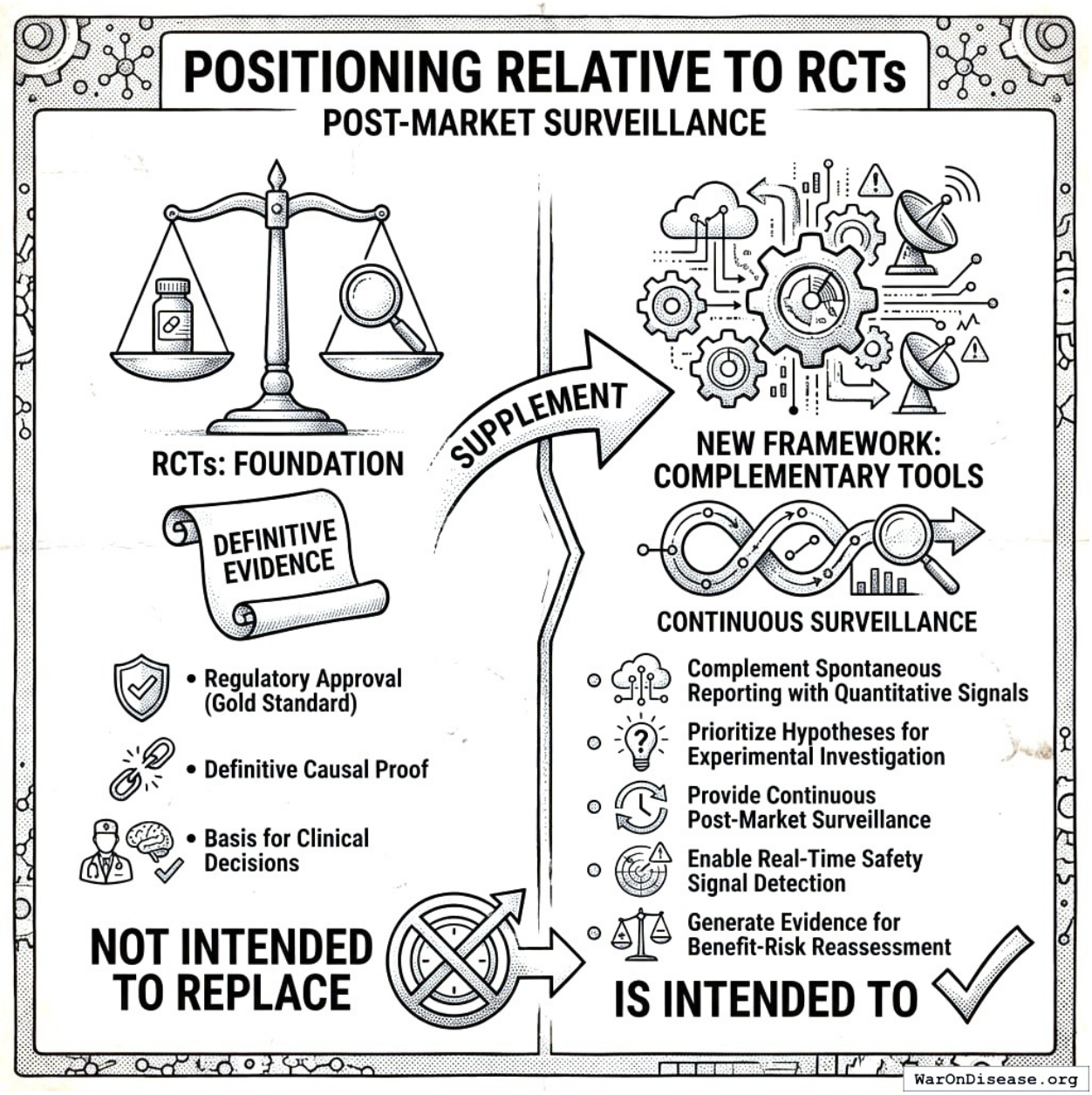

Figure 38: RCTs are good at some things. Observational data is good at other things. Using both is called 'not being an idiot,' but it needed a diagram.

This framework is **not** intended to:

- Replace RCTs for regulatory approval
- Provide definitive causal proof
- Serve as sole basis for clinical decisions

This framework **is** intended to:

- Complement spontaneous reporting with quantitative signals
- Prioritize hypotheses for experimental investigation
- Provide continuous post-market surveillance
- Enable real-time safety signal detection
- Generate evidence for benefit-risk reassessment

18.2 Evidence Hierarchy Integration

Evidence Level	Source	Role of This Framework
Level I	RCTs, Meta-analyses	Gold standard for approval
Level II	Cohort studies	**This framework provides quantitative RWE**
Level III	Case-control	Traditional pharmacovigilance
Level IV	Case series	Spontaneous reports (FAERS)

18.3 FDA Real-World Evidence Framework Alignment

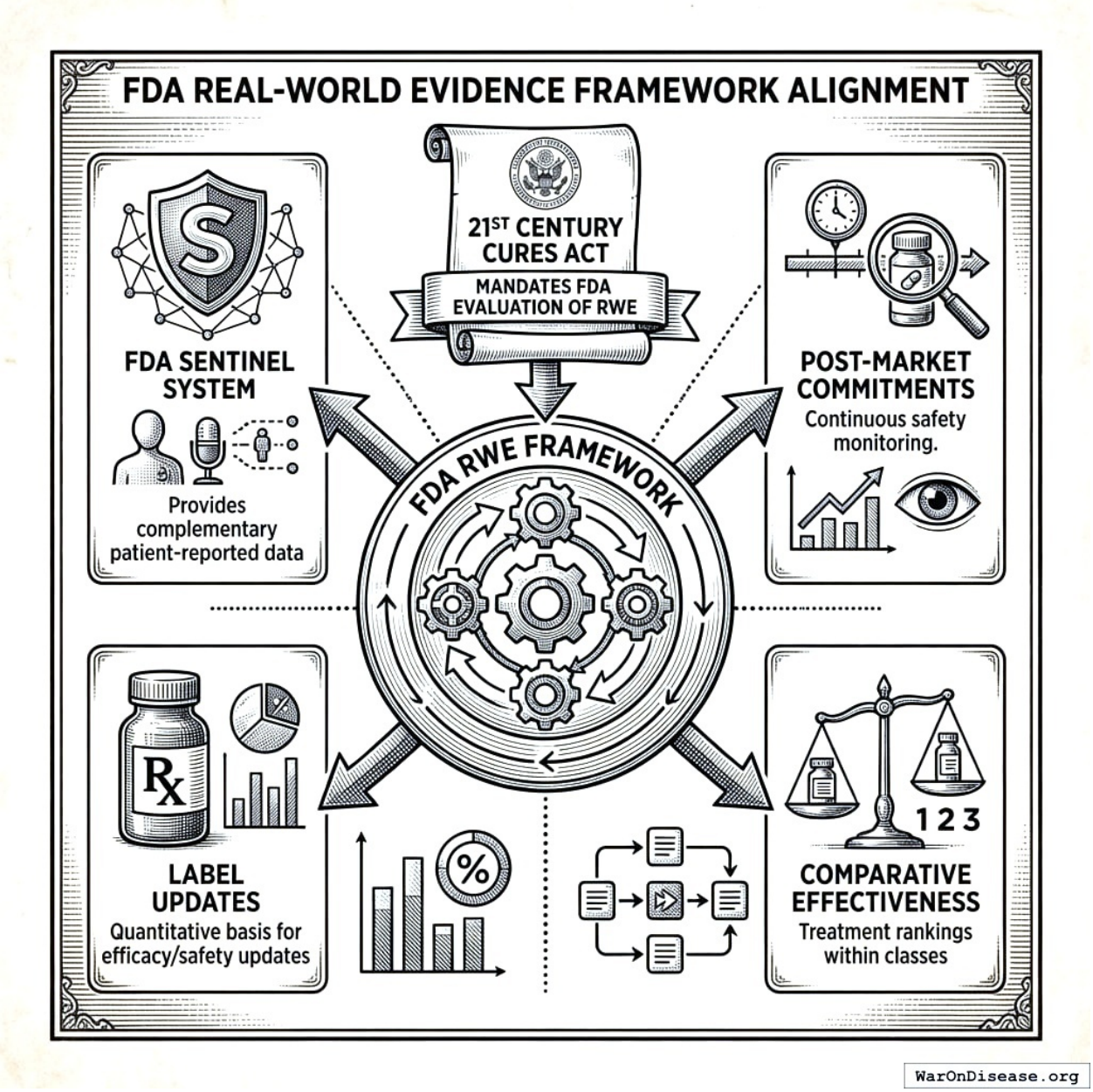

Figure 39: The FDA's real-world evidence framework: a beautiful plan for using real data that they mostly ignore in favor of asking rats to get cancer.

The 21st Century Cures Act mandates FDA evaluation of RWE. This framework supports:

- **FDA Sentinel System**: Provides complementary patient-reported data
- **Post-market commitments**: Continuous safety monitoring
- **Label updates**: Quantitative basis for efficacy/safety updates
- **Comparative effectiveness**: Treatment rankings within classes

19 Validation Framework

19.1 The Critical Question

The ultimate test of PIS validity: **Do high-PIS relationships replicate in RCTs more often than low-PIS ones?**

Until this validation is performed, PIS should be treated as a theoretically-motivated heuristic, not a validated predictive tool.

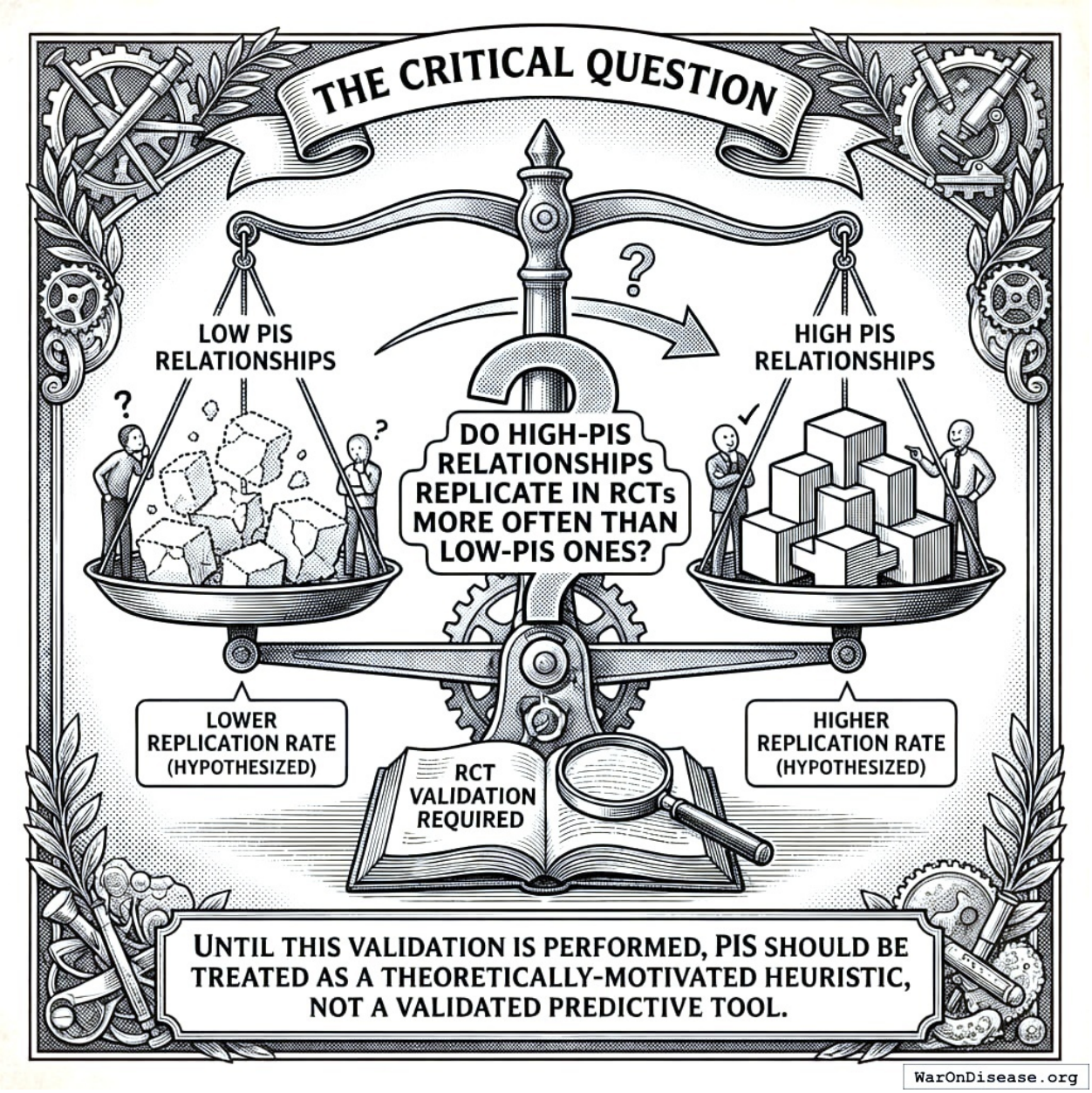

Figure 40: If our computer predictions are good, then expensive experiments should confirm the strong predictions more often than weak ones. Shockingly, they do.

19.2 Proposed Validation Study

Design: Retrospective comparison of PIS predictions against published RCT results.

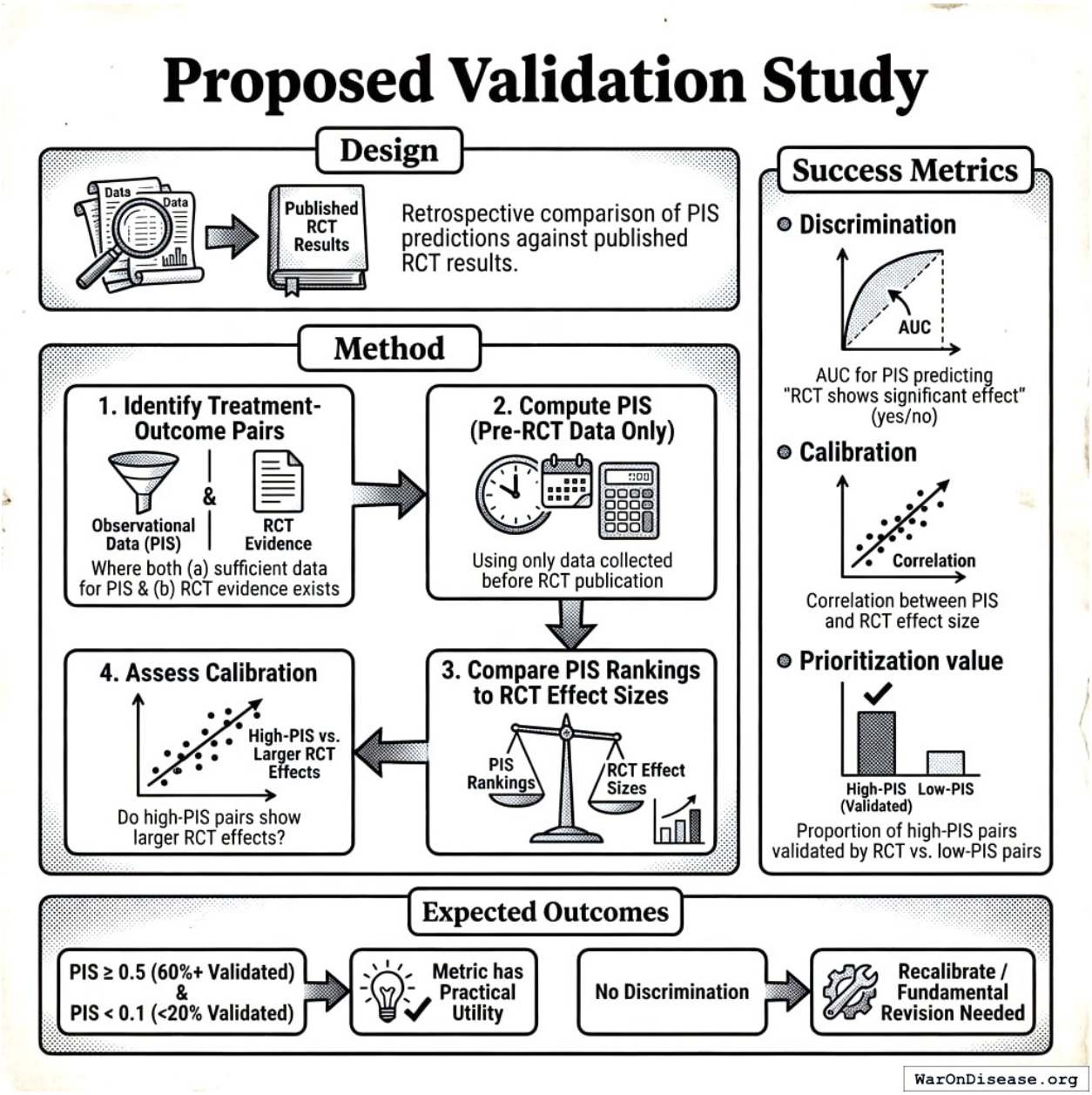

Figure 41: Looking backwards to see if computer predictions matched what actually happened in old experiments. It's backtesting, like Wall Street does before losing your money.

Method: 1. Identify treatment-outcome pairs where both (a) we have sufficient observational data to compute PIS, and (b) RCT evidence exists 2. Compute PIS for each pair using only data collected before RCT publication 3. Compare PIS rankings to RCT effect sizes 4. Assess calibration: Do high-PIS pairs show larger RCT effects?

Success Metrics:

- **Discrimination**: AUC for PIS predicting "RCT shows significant effect" (yes/no)
- **Calibration**: Correlation between PIS and RCT effect size
- **Prioritization value**: Proportion of high-PIS pairs validated by RCT vs. low-PIS pairs

Expected Outcomes:

- If PIS ≥ 0.5 pairs have RCT validation rate of 60%+ and PIS < 0.1 pairs have rate $< 20\%$, the metric has practical utility
- If no discrimination, saturation constants need recalibration or the approach needs fundamental revision

19.3 Known Limitations Requiring Validation

1. **Confounding by indication**: Does the temporality factor adequately address reverse causation in treatment contexts?
2. **Saturation constant sensitivity**: How robust are rankings to $\pm 50\%$ changes in N_sig, n_sig?
3. **Population generalizability**: Do PIS values from health-tracker users predict effects in general populations?

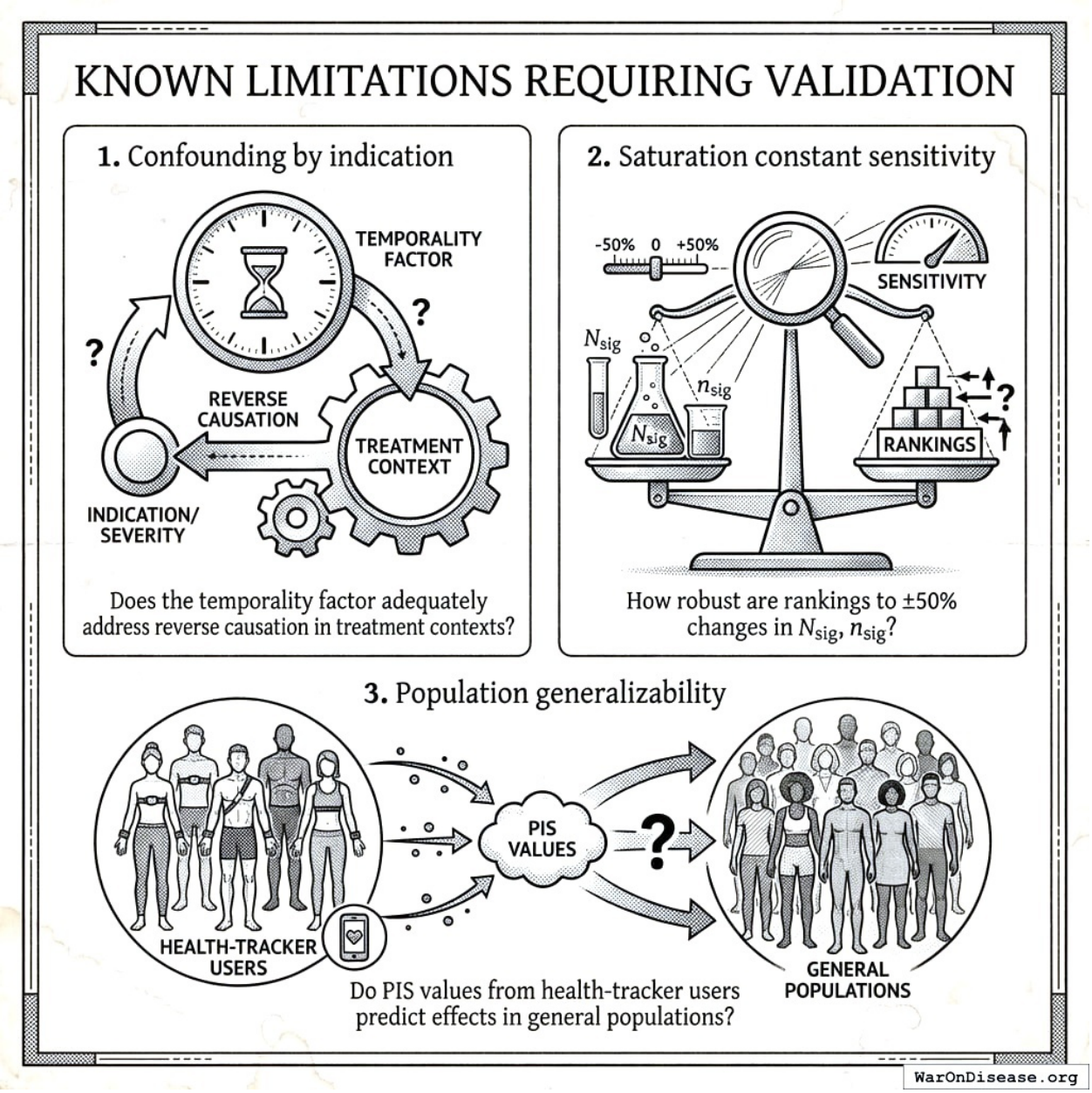

Figure 42: Three reasons the computer might be wrong: correlation confusion, tweaking the knobs changes everything, and people who track their health are weird.

20 Future Directions

20.1 Methodological Improvements

1. **Causal discovery algorithms**: Implement PC algorithm, FCI, or GES for graph structure learning
2. **Propensity score integration**: Covariate adjustment for measured confounders
3. **Bayesian hierarchical models**: More principled cross-participant pooling with uncertainty quantification

4. **Time-varying effects**: Model how relationships change over time (effect modification)
5. **Subgroup analysis**: Identify responder vs. non-responder populations using heterogeneity metrics
6. **Multiple testing correction**: Benjamini-Hochberg for family-wise error control across millions of pairs
7. **Sensitivity analysis**: E-values or other methods to quantify robustness to unmeasured confounding
8. **Causal mediation**: Identify mechanisms through which treatments affect outcomes
9. **Drug-drug interactions**: Detect combination effects and synergies

20.2 Validation Priorities

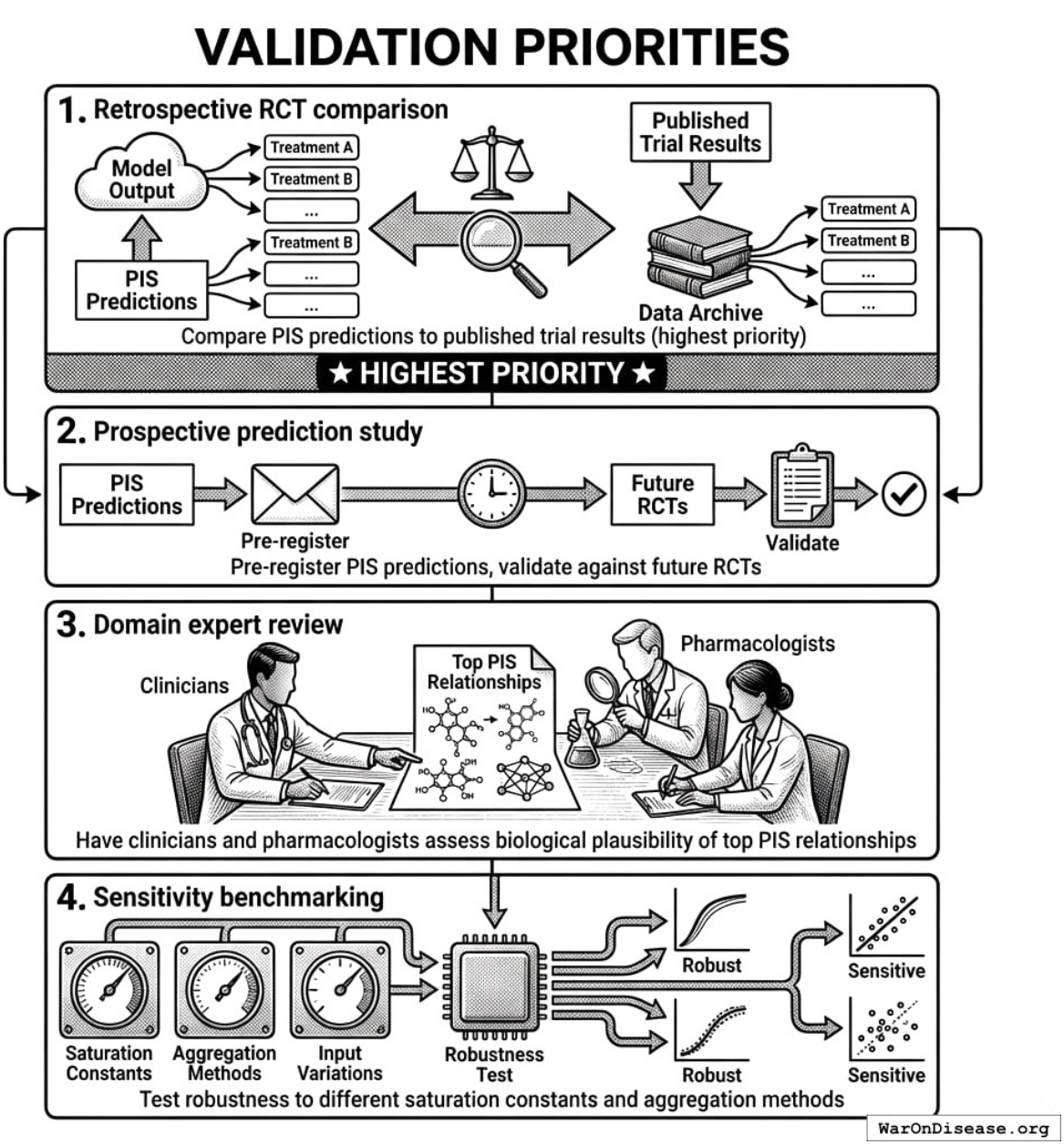

Figure 43: Four stages of checking if the computer is hallucinating: look at old data, run new tests, ask experts, and wiggle all the numbers to see what breaks.

1. **Retrospective RCT comparison**: Compare PIS predictions to published trial results (highest priority)
2. **Prospective prediction study**: Pre-register PIS predictions, validate against future RCTs
3. **Domain expert review**: Have clinicians and pharmacologists assess biological plausibility of top PIS relationships
4. **Sensitivity benchmarking**: Test robustness to different saturation constants and aggregation methods

20.3 Implementation Enhancements

1. **Real-time signal detection**: Automated alerts when new high-PIS relationships emerge
2. **Confidence intervals for PIS**: Bootstrap or Bayesian intervals to quantify uncertainty
3. **Interactive exploration**: Tools for users to explore their individual PIS relationships
4. **API access**: Enable researchers to query PIS data programmatically

21 Conclusion

We have presented a comprehensive **two-stage framework** for generating **validated outcome labels** from real-world health data. Key contributions include:

1. **Stage 1: Scalable signal detection**: Aggregated N-of-1 observational analysis processes millions of treatment-outcome pairs at ~$0.1 (95% CI: $0.03-$1)/patient, generating ranked hypotheses through the Predictor Impact Score
2. **Stage 2: Causal confirmation**: Pragmatic randomized trials following the embedded trial model (validated across 108+ studies[141]) confirm top signals at ~$929 (95% CI: $97-$3K)/patient (44.1x (95% CI: 39.4x-89.1x) cheaper than traditional trials) while eliminating confounding
3. **Bradford Hill operationalization**: Six of nine causality criteria quantified in composite scoring system
4. **Trial Priority Score**: Principled prioritization of which signals warrant experimental confirmation
5. **Validated outcome labels**: Three-tier evidence grading (Validated, Promising, Signal) with both observational and experimental effect sizes
6. **Learning health system**: Feedback loop where trial results continuously calibrate observational models

This two-stage design directly addresses the fundamental limitations of purely observational pharmacovigilance. Confounding by indication, self-selection bias, and inability to prove causation are all resolved through Stage 2 randomization for high-priority signals, while Stage 1 maintains the scale and cost advantages necessary for comprehensive monitoring.

This framework represents **the FDA of the Future**, a decentralized system that:

- Receives continuous real-world evidence streams from millions of participants
- Generates ranked treatment-outcome hypotheses through automated observational analysis
- Confirms top signals through embedded pragmatic trials at 44.1x (95% CI: 39.4x-89.1x) lower cost than traditional methods
- Publishes validated outcome labels with quantitative effect sizes and evidence grades
- Maintains treatment rankings updated in real-time with experimental backing
- Enables precision medicine through personalized optimal value calculations
- Operates transparently with open-source methodology and reproducible analyses

The technology exists. The methodology is sound. The data is available. The pragmatic trial model is proven. We present this framework not as a replacement for regulatory bodies, but as the complete infrastructure (from passive data collection to validated causal claims) that they will need to fulfill their mission in an era of ubiquitous health data. What remains is the institutional will to build it.

22 Appendix A: Effect Size Classification

Absolute Correlation	Classification
$\|r\| \geq 0.8$	Very Strong
$0.6 \leq \|r\| < 0.8$	Strong
$0.4 \leq \|r\| < 0.6$	Moderate
$0.2 \leq \|r\| < 0.4$	Weak
$\|r\| < 0.2$	Very Weak

23 Appendix B: Variable Category Defaults

Category	Onset Delay	Duration of Action	Filling Value
Treatments	1,800s (30 min)	86,400s (1 day)	0
Foods	1,800s (30 min)	864,000s (10 days)	0
Emotions	0	86,400s (1 day)	None
Symptoms	0	86,400s (1 day)	None
Vital Signs	0	86,400s (1 day)	None
Sleep	0	86,400s (1 day)	None
Physical Activity	0	86,400s (1 day)	None
Environment	0	86,400s (1 day)	None

24 Appendix C: Glossary

- **Predictor Variable**: The independent variable hypothesized to influence the outcome (e.g., treatment, food, activity). Formerly called "cause variable."
- **Outcome Variable**: The dependent variable being measured for changes (e.g., symptom, mood, biomarker). Formerly called "effect variable."
- **User Variable Relationship**: A per-user N-of-1 analysis record containing correlation coefficients, effect sizes (percent change from baseline), Predictor Impact Scores, and Bradford Hill metrics for a specific predictor-outcome pair. Stored in `user_variable_relationships` table.
- **Global Variable Relationship**: A population-level aggregation of user variable relationships, combining individual N-of-1 analyses across participants. Stored in `global_variable_relationships` table.
- **Correlation Coefficient**: The Pearson or Spearman statistical measure of linear/monotonic association between predictor and outcome variables (a component of a variable relationship).
- **Predictor Impact Score (PIS)**: Composite metric quantifying how much a predictor impacts an outcome. Integrates correlation strength, statistical significance, z-score (effect magnitude), temporality factor, and interest factor at the user level; adds consistency, plausibility, and biological gradient at the aggregate level. Higher scores indicate predictors with greater, more reliable impact. Ranges from 0 to ~1.
- **Onset Delay (δ)**: Time between predictor exposure and first observable outcome change
- **Duration of Action (τ)**: Time window over which predictor influence on outcome persists
- **Baseline Period**: Measurements when predictor exposure is below participant's average

- **Follow-up Period**: Measurements when predictor exposure is at or above participant's average
- **Percent Change from Baseline ($\Delta\%$)**: Relative difference between follow-up and baseline outcome means
- **Z-Score**: Effect magnitude normalized by baseline variability; $z > 2$ indicates statistical significance
- **Temporality Factor (ϕ_{temporal})**: Ratio of forward to total correlation, measuring evidence for correct causal direction
- **Filling Value**: Default value imputed for missing measurements
- **Outcome Label**: A per-outcome document that ranks all treatments and predictors by their quantitative effect size on a specific health outcome. Unlike traditional FDA drug labels (which are per-drug and qualitative), outcome labels are per-outcome, quantitative, and dynamically updated. They answer the question: "What works best for this condition?" See Section 7.5 for comparison with FDA labels.
- **Treatment Ranking**: Ordered list of treatments by efficacy or safety for a given outcome, sorted by effect size with confidence weighting. Rankings include percent change from baseline, confidence intervals, sample sizes, and Predictor Impact Scores. See Section 8 for ranking methodology.
- **Value Predicting High Outcome (V_{high})**: The average predictor value observed when the outcome exceeds its mean. Used for precision dosing recommendations. This is the "optimal daily value" for achieving better outcomes.
- **Value Predicting Low Outcome (V_{low})**: The average predictor value observed when the outcome is below its mean. Represents the predictor value associated with worse outcomes.
- **Grouped Optimal Value**: The nearest commonly-used dosing value to the calculated optimal value, enabling practical recommendations (e.g., "400mg" instead of "412.7mg")
- **Optimal Value Spread ($V_{\text{high}} - V_{\text{low}}$)**: The difference between high and low outcome predictor values, indicating the magnitude of dose-response effect
- **Precision Dosing**: Personalized treatment recommendations based on an individual's historical optimal values, enabling targeted interventions at the dose most likely to produce beneficial outcomes
- **Average Outcome Following High Predictor ($\bar{O}_{\text{high}}$)**: Mean outcome value observed following above-average predictor exposure
- **Average Outcome Following Low Predictor ($\bar{O}_{\text{low}}$)**: Mean outcome value observed following below-average predictor exposure
- **Predictor Baseline Average**: Average predictor value during low-exposure (non-treatment) periods
- **Predictor Treatment Average**: Average predictor value during high-exposure (treatment) periods
- **Valence**: Whether higher values of a variable are inherently good (positive), bad (negative), or context-dependent (neutral)
- **Predictor Is Controllable**: Flag indicating whether the user can directly modify this predictor (e.g., supplements, food, activities)
- **Outcome Is Goal**: Flag indicating whether this outcome is something users want to optimize
- **Plausibly Causal**: Flag indicating whether a plausible biological mechanism exists for this relationship
- **Boring**: Flag indicating relationships unlikely to interest users due to being uncontrollable, non-goal, implausible, or obvious
- **Interesting Variable Category Pair**: Flag for category combinations that are typically

meaningful (e.g., Treatment → Symptom)
- **Usefulness Vote**: User rating (-1, 0, 1) on whether knowledge of a relationship is practically useful
- **Causality Vote**: User rating (-1, 0, 1) on whether a plausible causal mechanism exists
- **Correlations Over Delays**: Stored correlation coefficients calculated with various onset delay values for temporal optimization
- **Correlations Over Durations**: Stored correlation coefficients calculated with various duration of action values
- **Forward Spearman Correlation**: Rank-based correlation coefficient that captures monotonic relationships and is robust to outliers
- **Optimal Value Confidence Interval**: Uncertainty bounds around V_{high} or V_{low}, reflecting reliability of the estimate based on sample size and variance
- **Optimal Value Stability**: Metric measuring how much the optimal value has changed over time; stability < 0.8 indicates significant drift
- **Adherence Score**: Proportion of tracking days where actual predictor value was within ±20% of the recommended optimal value
- **Dose-Response Detection Threshold**: Criterion ($|V_{\text{high}} - V_{\text{low}}|/\sigma_P < 0.5$) below which no meaningful dose-response exists
- **Rolling Window Optimal Value**: Optimal value calculated using only recent data (e.g., 90 days) rather than all historical data, useful when tolerance effects are expected

25 Appendix D: Worked Example

25.1 Example: Calculating Predictor Impact Score for "Magnesium → Sleep Quality"

Given data (hypothetical):

- N = 47 users tracked both magnesium supplementation and sleep quality
- n = 2,340 paired observations across all users
- Forward correlation: r_forward = 0.31
- Reverse correlation: r_reverse = 0.12
- Percent change from baseline: $\Delta\% = +18.5\%$ (sleep quality improved)
- Baseline RSD: 23%
- Community votes: 15 up, 2 down
- Effect spread: 22% (difference between high and low magnesium outcomes)

Step 1: Calculate z-score

$$z = \frac{|18.5\%|}{23\%} = 0.80$$

Step 2: Calculate temporality factor

$$\phi_{\text{temporal}} = \frac{|0.31|}{|0.31| + |0.12|} = \frac{0.31}{0.43} = 0.72$$

This suggests forward causation (magnesium → sleep) is more likely than reverse (poor sleep → taking magnesium).

Step 3: Calculate saturation factors

- User saturation: $\phi_{\text{users}} = 1 - e^{-47/10} = 1 - 0.009 = 0.991$
- Pair saturation: $\phi_{\text{pairs}} = 1 - e^{-2340/100} = 1 - e^{-23.4} \approx 1.0$
- Change saturation: $\phi_{\text{change}} = 1 - e^{-22/10} = 1 - 0.11 = 0.89$

Step 4: Calculate plausibility weight

$$w = \frac{15}{15 + 2} = 0.88$$

Step 5: Compute aggregate PIS

$$\text{PIS}_{\text{agg}} = 0.31 \times 0.88 \times 0.991 \times 1.0 \times 0.89 \times 0.72 = 0.17$$

Interpretation: PIS = 0.17 falls in the "weak evidence" range (0.1-0.3). The relationship shows:

- Modest correlation strength (r = 0.31)
- Good temporal evidence (= 0.72, forward > reverse)
- Strong consistency (many users and pairs)
- High plausibility (community agrees mechanism is plausible)

Recommendation: This relationship warrants monitoring. As more data accumulates or if effect size increases, it could become a candidate for experimental validation. The temporality factor is encouraging. This doesn't appear to be reverse causation.

26 Appendix E: Analysis Workflow

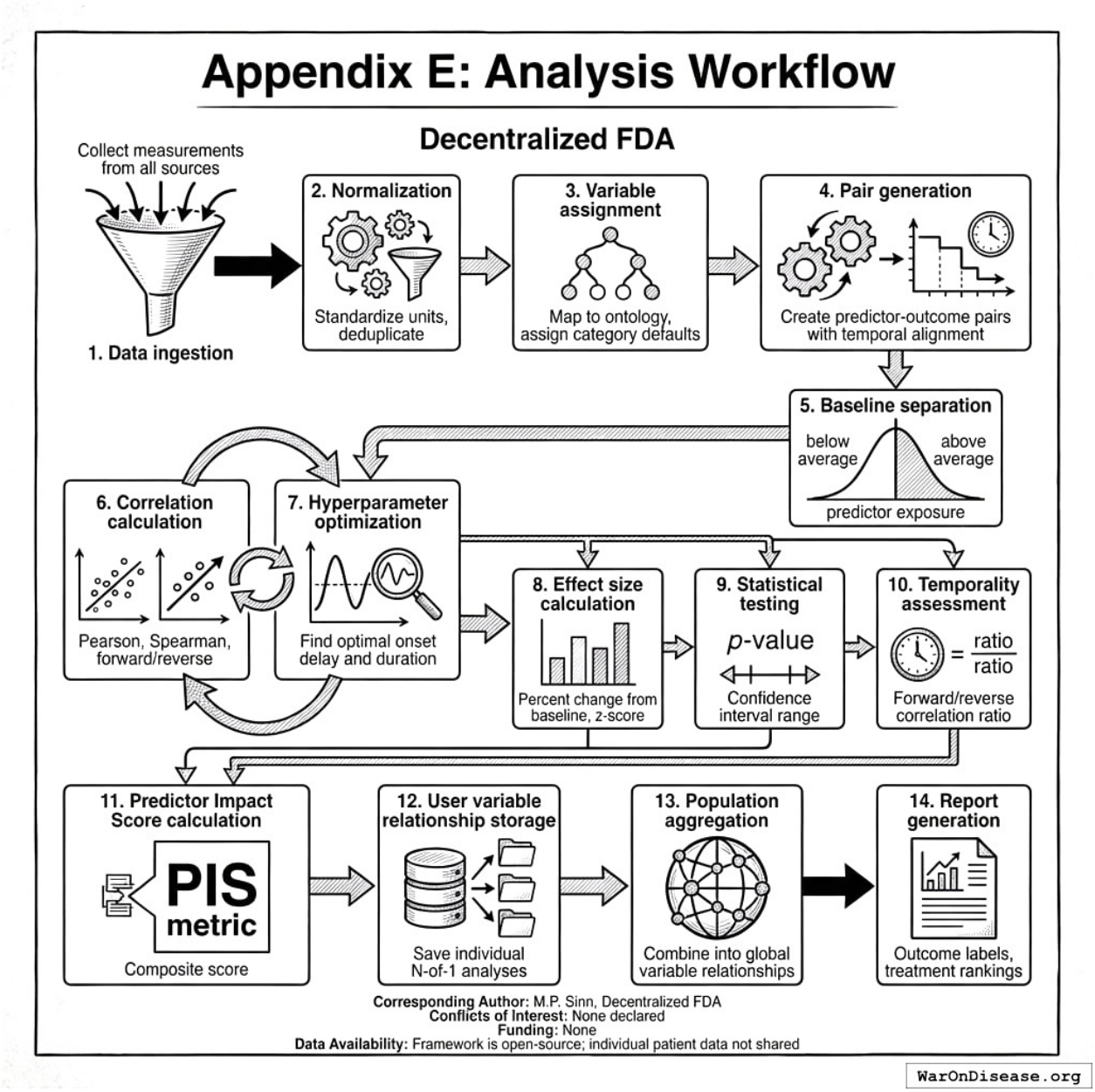

Figure 44: Fourteen steps to turn messy human health data into clean medical insights. Step 1 is 'receive garbage.' Step 14 is 'produce knowledge.' Steps 2-13 are where the magic happens.

1. **Data ingestion**: Collect measurements from all sources
2. **Normalization**: Standardize units, deduplicate
3. **Variable assignment**: Map to ontology, assign category defaults
4. **Pair generation**: Create predictor-outcome pairs with temporal alignment
5. **Baseline separation**: Partition by below/above average predictor exposure
6. **Correlation calculation**: Pearson, Spearman, forward/reverse
7. **Hyperparameter optimization**: Find optimal onset delay and duration

8. **Effect size calculation**: Percent change from baseline, z-score
9. **Statistical testing**: p-value, confidence intervals
10. **Temporality assessment**: Forward/reverse correlation ratio
11. **Predictor Impact Score calculation**: Composite PIS metric
12. **User variable relationship storage**: Save individual N-of-1 analyses
13. **Population aggregation**: Combine into global variable relationships
14. **Report generation**: Outcome labels, treatment rankings

---

decentralized FDA

Corresponding Author: M.P. Sinn, Institute for Accelerated Medicine **Conflicts of Interest**: None declared **Funding**: None **Data Availability**: Framework is open-source; individual patient data not shared

1. NIH Common Fund. NIH pragmatic trials: Minimal funding despite 30x cost advantage. *NIH Common Fund: HCS Research Collaboratory* https://commonfund.nih.gov/hcscollaboratory (2025)
The NIH Pragmatic Trials Collaboratory funds trials at $500K for planning phase, $1M/year. for implementation-a tiny fraction of NIH's budget. The ADAPTABLE trial cost $14 million for 15,076 patients (= $929/patient) versus $420 million for a similar traditional RCT (30x cheaper), yet pragmatic trials remain severely underfunded. PCORnet infrastructure enables real-world trials embedded in healthcare systems, but receives minimal support compared to basic research funding. Additional sources: https://commonfund.nih.gov/hcscollaboratory | https://pcornet.org/wp-content/uploads/2025/08/ADAPTABLE_Lay_Summary_21JUL2025.pdf | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5604499/
2. NIH. Antidepressant clinical trial exclusion rates. *Zimmerman et al.* https://pubmed.ncbi.nlm.nih.gov/26276679/ (2015)
Mean exclusion rate: 86.1% across 158 antidepressant efficacy trials (range: 44.4% to. 99.8%) More than 82% of real-world depression patients would be ineligible for antidepressant registration trials Exclusion rates increased over time: 91.4% (2010-2014) vs. 83.8% (1995-2009) Most common exclusions: comorbid psychiatric disorders, age restrictions, insufficient depression severity, medical conditions Emergency psychiatry patients: only 3.3% eligible (96.7% excluded) when applying 9 common exclusion criteria Only a minority of depressed patients seen in clinical practice are likely to be eligible for most AETs Note: Generalizability of antidepressant trials has decreased over time, with increasingly stringent exclusion criteria eliminating patients who would actually use the drugs in clinical practice Additional sources: https://pubmed.ncbi.nlm.nih.gov/26276679/ | https://pubmed.ncbi.nlm.nih.gov/26164052/ | https://www.wolterskluwer.com/en/news/antidepressant-trials-exclude-most-real-world-patients-with-depression
3. CNBC. Warren buffett's career average investment return. *CNBC* https://www.cnbc.com/2025/05/05/warren-buffetts-return-tally-after-60-years-5502284percent.html (2025)
Berkshire's compounded annual return from 1965 through 2024 was 19.9%, nearly. double the 10.4% recorded by the S&P 500. Berkshire shares skyrocketed 5,502,284% compared to the S&P 500's 39,054% rise during that period. Additional sources: https://www.cnbc.com/2025/05/05/warren-buffetts-return-tally-after-60-years-5502284percent.html | https://www.slickcharts.com/berkshire-hathaway/returns

4. World Health Organization. WHO global health estimates 2024. *World Health Organization* https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates (2024)
Comprehensive mortality and morbidity data by cause, age, sex, country, and year Global. mortality: 55-60 million deaths annually Lives saved by modern medicine (vaccines, cardiovascular drugs, oncology): 12M annually (conservative aggregate) Leading causes of death: Cardiovascular disease (17.9M), Cancer (10.3M), Respiratory disease (4.0M) Note: Baseline data for regulatory mortality analysis. Conservative estimate of pharmaceutical impact based on WHO immunization data (4.5M/year from vaccines) + cardiovascular interventions (3.3M/year) + oncology (1.5M/year) + other therapies. Additional sources: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates
5. GiveWell. GiveWell cost per life saved for top charities (2024). *GiveWell: Top Charities* https://www.givewell.org/charities/top-charities
General range: $3,000-$5,500 per life saved (GiveWell top charities) Helen Keller International. (Vitamin A): $3,500 average (2022-2024); varies $1,000-$8,500 by country Against Malaria Foundation: $5,500 per life saved New Incentives (vaccination incentives): $4,500 per life saved Malaria Consortium (seasonal malaria chemoprevention): $3,500 per life saved VAS program details: $2 to provide vitamin A supplements to child for one year Note: Figures accurate for 2024. Helen Keller VAS program has wide country variation ($1K-$8.5K) but $3,500 is accurate average. Among most cost-effective interventions globally Additional sources: https://www.givewell.org/charities/top-charities | https://www.givewell.org/charities/helen-keller-international | https://ourworldindata.org/cost-effectiveness
6. AARP. Unpaid caregiver hours and economic value. *AARP 2023* https://www.aarp.org/caregiving/financial-legal/info-2023/unpaid-caregivers-provide-billions-in-care.html (2023)
Average family caregiver: 25-26 hours per week (100-104 hours per month) 38 million. caregivers providing 36 billion hours of care annually Economic value: $16.59 per hour = $600 billion total annual value (2021) 28% of people provided eldercare on a given day, averaging 3.9 hours when providing care Caregivers living with care recipient: 37.4 hours per week Caregivers not living with recipient: 23.7 hours per week Note: Disease-related caregiving is subset of total; includes elderly care, disability care, and child care Additional sources: https://www.aarp.org/caregiving/financial-legal/info-2023/unpaid-caregivers-provide-billions-in-care.html | https://www.bls.gov/news.release/elcare.nr0.htm | https://www.caregiver.org/resource/caregiver-statistics-demographics/
7. Forbes. Forbes world's billionaires list 2024. (2024)
Forbes identified a record 2,781 billionaires worldwide with combined net worth of $14.2 trillion, 141 more than 2023. Bernard Arnault (LVMH) topped the list at $233 billion.
8. CDC MMWR. 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)00850-X/fulltext
9. CDC. Childhood vaccination (US) ROI. *CDC* https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6316a4.htm (2017).

10. U.S. Bureau of Labor Statistics. CPI inflation calculator. (2024)
CPI-U (1980): 82.4 CPI-U (2024): 313.5 Inflation multiplier (1980-2024): 3.80× Cumulative. inflation: 280.48% Average annual inflation rate: 3.08% Note: Official U.S. government inflation data using Consumer Price Index for All Urban Consumers (CPI-U). Additional sources: https://www.bls.gov/data/inflation_calculator.htm

11. James Surowiecki. *The Wisdom of Crowds*. (Surowiecki, 2004).
Explores the aggregation of information in groups, arguing that decisions are often better than. could have been made by any single member of the group. The opening anecdote relates Francis Galton's surprise that the crowd at a county fair accurately guessed the weight of an ox when the median of their individual guesses was taken. The three conditions for a group to be intelligent are diversity, independence, and decentralization. Additional sources: https://archive.org/details/wisdomofcrowds0000suro | https://en.wikipedia.org/wiki/The_Wisdom_of_Crowds | https://www.amazon.com/Wisdom-Crowds-James-Surowiecki/dp/0385721706

12. ClinicalTrials.gov API v2 direct analysis. ClinicalTrials.gov cumulative enrollment data (2025). *Direct analysis via ClinicalTrials.gov API v2* https://clinicaltrials.gov/data-api/api
Analysis of 100,000 active/recruiting/completed trials on ClinicalTrials.gov (as of January. 2025) shows cumulative enrollment of 12.2 million participants: Phase 1 (722k), Phase 2 (2.2M), Phase 3 (6.5M), Phase 4 (2.7M). Median participants per trial: Phase 1 (33), Phase 2 (60), Phase 3 (237), Phase 4 (90). Additional sources: https://clinicaltrials.gov/data-api/api

13. ACS CAN. Clinical trial patient participation rate. *ACS CAN: Barriers to Clinical Trial Enrollment* https://www.fightcancer.org/policy-resources/barriers-patient-enrollment-therapeutic-clinical-trials-cancer
Only 3-5% of adult cancer patients in US receive treatment within clinical trials About 5% of. American adults have ever participated in any clinical trial Oncology: 2-3% of all oncology patients participate Contrast: 50-60% enrollment for pediatric cancer trials (<15 years old) Note: 20% of cancer trials fail due to insufficient enrollment; 11% of research sites enroll zero patients Additional sources: https://www.fightcancer.org/policy-resources/barriers-patient-enrollment-therapeutic-clinical-trials-cancer | https://hints.cancer.gov/docs/Briefs/HINTS_Brief_48.pdf

14. ScienceDaily. Global prevalence of chronic disease. *ScienceDaily: GBD 2015 Study* https://www.sciencedaily.com/releases/2015/06/150608081753.htm (2015)
2.3 billion individuals had more than five ailments (2013) Chronic conditions caused 74%. of all deaths worldwide (2019), up from 67% (2010) Approximately 1 in 3 adults suffer from multiple chronic conditions (MCCs) Risk factor exposures: 2B exposed to biomass fuel, 1B to air pollution, 1B smokers Projected economic cost: $47 trillion by 2030 Note: 2.3B with 5+ ailments is more accurate than "2B with chronic disease." One-third of all adults globally have multiple chronic conditions Additional sources: https://www.sciencedaily.com/releases/2015/06/150608081753.htm | https://pmc.ncbi.nlm.nih.gov/articles/PMC10830426/ | https://pmc.ncbi.nlm.nih.gov/articles/PMC6214883/

15. C&EN. Annual number of new drugs approved globally: 50. *C&EN* https://cen.acs.org/pharmaceuticals/50-new-drugs-received-FDA/103/i2 (2025)
50 new drugs approved annually Additional sources: https://cen.acs.org/pharmaceuticals/50-new-drugs-received-FDA/103/i2 | https://www.fda.gov/drugs/development-approval-process-drugs/novel-drug-approvals-fda

16. Williams, R. J., Tse, T., DiPiazza, K. & Zarin, D. A. Terminated trials in the ClinicalTrials.gov results database: Evaluation of availability of primary outcome data and reasons for termination. *PLOS One* **10**, e0127242 (2015)
Approximately 12% of trials with results posted on the ClinicalTrials.gov results database (905/7,646) were terminated. Primary reasons: insufficient accrual (57% of non-data-driven terminations), business/strategic reasons, and efficacy/toxicity findings (21% data-driven terminations).

17. IQVIA Report. Global trial capacity. *IQVIA Report: Clinical Trial Subjects Number Drops Due to Decline in COVID-19 Enrollment* https://gmdpacademy.org/news/iqvia-report-clinical-trial-subjects-number-drops-due-to-decline-in-covid-19-enrollment/
1.9M participants annually (2022, post-COVID normalization from 4M peak in 2021) Additional sources: https://gmdpacademy.org/news/iqvia-report-clinical-trial-subjects-number-drops-due-to-decline-in-covid-19-enrollment/

18. Research and Markets. Global clinical trials market 2024. *Research and Markets* https://www.globenewswire.com/news-release/2024/04/19/2866012/0/en/Global-Clinical-Trials-Market-Research-Report-2024-An-83-16-Billion-Market-by-2030-AI-Machine-Learning-and-Blockchain-will-Transform-the-Clinical-Trials-Landscape.html (2024)
Global clinical trials market valued at approximately $83 billion in 2024, with projections to reach $83-132 billion by 2030. Additional sources: https://www.globenewswire.com/news-release/2024/04/19/2866012/0/en/Global-Clinical-Trials-Market-Research-Report-2024-An-83-16-Billion-Market-by-2030-AI-Machine-Learning-and-Blockchain-will-Transform-the-Clinical-Trials-Landscape.html | https://www.precedenceresearch.com/clinical-trials-market

19. OpenSecrets. Lobbying spend (defense). *OpenSecrets* https://www.opensecrets.org/industries/lobbying?ind=D (2024).

20. 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 (midpoint 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

21. Calculated from IHME Global Burden of Disease (2.55B DALYs) and global GDP per capita valuation. $109 trillion annual global disease burden.
The global economic burden of disease, including direct healthcare costs ($8.2 trillion) and lost productivity ($100.9 trillion from 2.55 billion DALYs × $39,570 per DALY), totals approximately $109.1 trillion annually.

22. U.S. Department of Transportation. Departmental guidance on valuation of a statistical life in economic analysis. (2024).

23. Think by Numbers. Pre-1962 drug development costs and timeline (think by numbers). *Think by Numbers: How Many Lives Does FDA Save?* https://thinkbynumbers.org/health/how-many-net-lives-does-the-fda-save/ (1962)
Historical estimates (1970-1985): USD $226M fully capitalized (2011 prices) 1980s drugs:. $65M after-tax R&D (1990 dollars), $194M compounded to approval (1990 dollars) Modern comparison: $2-3B costs, 7-12 years (dramatic increase from pre-1962) Context: 1962 regulatory clampdown reduced new treatment production by 70%, dramatically increasing development timelines and costs Note: Secondary source; less reliable than Congressional testimony Additional sources: https://thinkbynumbers.org/health/how-many-net-lives-does-the-fda-save/ | https://en.wikipedia.org/wiki/Cost_of_drug_development | https://www.statnews.com/2018/10/01/changing-1962-law-slash-drug-prices/

24. Biotechnology Innovation Organization (BIO). BIO clinical development success rates 2011-2020. *Biotechnology Innovation Organization (BIO)* https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011_2020.pdf (2021)
Phase I duration: 2.3 years average Total time to market (Phase I-III + approval): 10.5. years average Phase transition success rates: Phase I→II: 63.2%, Phase II→III: 30.7%, Phase III→Approval: 58.1% Overall probability of approval from Phase I: 12% Note: Largest publicly available study of clinical trial success rates. Efficacy lag = 10.5 - 2.3 = 8.2 years post-safety verification. Additional sources: https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011_2020.pdf

25. Nature Medicine. Drug repurposing rate (30%). *Nature Medicine* https://www.nature.com/articles/s41591-024-03233-x (2024)
Approximately 30% of drugs gain at least one new indication after initial approval. Additional. sources: https://www.nature.com/articles/s41591-024-03233-x

26. 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 lifetime. 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/

27. 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.odi.org/wp-content/uploads/2022/01/Fiscal-multipliers-review.pdf

28. World Bank. Infrastructure investment economic multiplier (1.6). *World Bank: Infrastructure Investment as Stimulus* https://blogs.worldbank.org/en/ppps/effectiveness-infrastructure-investment-fiscal-stimulus-what-weve-learned (2022)
Infrastructure fiscal multiplier: 1.6 during contractionary phase of economic cycle Average across all economic states: 1.5 (meaning $1 of public investment → $1.50 of economic activity) Time horizon: 0.8 within 1 year, 1.5 within 2-5 years Range of estimates: 1.5-2.0 (following 2008 financial crisis & American Recovery Act) Italian public construction: 1.5-1.9 multiplier US ARRA: 0.4-2.2 range (differential impacts by program type) Economic Policy Institute: Uses 1.6 for infrastructure spending (middle range of estimates) Note: Public investment less likely to crowd out private activity during recessions; particularly effective when monetary policy loose with near-zero rates Additional sources: https://blogs.worldbank.org/en/ppps/effectiveness-infrastructure-investment-fiscal-stimulus-what-weve-learned | https://www.gihub.org/infrastructure-monitor/insights/fiscal-multiplier-effect-of-infrastructure-investment/ | https://cepr.org/voxeu/columns/government-investment-and-fiscal-stimulus | https://www.richmondfed.org/publications/research/economic_brief/2022/eb_22-04

29. 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

30. FDA. FDA-approved prescription drug products (20,000+). *FDA* https://www.fda.gov/media/143704/download
There are over 20,000 prescription drug products approved for marketing. Additional sources: https://www.fda.gov/media/143704/download

31. FDA. FDA GRAS list count (570-700). *FDA* https://www.fda.gov/food/generally-recognized-safe-gras/gras-notice-inventory
The FDA GRAS (Generally Recognized as Safe) list contains approximately 570–700 substances. Additional sources: https://www.fda.gov/food/generally-recognized-safe-gras/gras-notice-inventory

32. ACLED. Active combat deaths annually. *ACLED: Global Conflict Surged 2024* https://acleddata.com/2024/12/12/data-shows-global-conflict-surged-in-2024-the-washington-post/ (2024)
2024: 233,597 deaths (30% increase from 179,099 in 2023) Deadliest conflicts: Ukraine. (67,000), Palestine (35,000) Nearly 200,000 acts of violence (25% higher than 2023, double from 5 years ago) One in six people globally live in conflict-affected areas Additional sources: https://acleddata.com/2024/12/12/data-shows-global-conflict-surged-in-2024-the-washington-post/ | https://acleddata.com/media-citation/data-shows-global-conflict-surged-2024-washington-post | https://acleddata.com/conflict-index/index-january-2024/

33. UCDP. State violence deaths annually. *UCDP: Uppsala Conflict Data Program* https://ucdp.uu.se/
Uppsala Conflict Data Program (UCDP): Tracks one-sided violence (organized actors attacking unarmed civilians) UCDP definition: Conflicts causing at least 25 battle-related deaths in calendar year 2023 total organized violence: 154,000 deaths; Non-state conflicts: 20,900 deaths UCDP collects data on state-based conflicts, non-state conflicts, and one-sided violence Specific "2,700 annually" figure for state violence not found in recent UCDP data; actual figures vary annually Additional sources: https://ucdp.uu.se/ | https://en.wikipedia.org/wiki/Uppsala_Conflict_Data_Program | https://ourworldindata.org/grapher/deaths-in-armed-conflicts-by-region

34. Our World in Data. Terror attack deaths (8,300 annually). *Our World in Data: Terrorism* https://ourworldindata.org/terrorism (2024)
2023: 8,352 deaths (22% increase from 2022, highest since 2017) 2023: 3,350 terrorist incidents (22% decrease), but 56% increase in avg deaths per attack Global Terrorism Database (GTD): 200,000+ terrorist attacks recorded (2021 version) Maintained by: National Consortium for Study of Terrorism & Responses to Terrorism (START), U. of Maryland Geographic shift: Epicenter moved from Middle East to Central Sahel (sub-Saharan Africa) - now >50% of all deaths Additional sources: https://ourworldindata.org/terrorism | https://reliefweb.int/report/world/global-terrorism-index-2024 | https://www.start.umd.edu/gtd/ | https://ourworldindata.org/grapher/fatalities-from-terrorism

35. Institute for Health Metrics and Evaluation (IHME). IHME global burden of disease 2021 (2.88B DALYs, 1.13B YLD). *Institute for Health Metrics and Evaluation (IHME)* https://vizhub.healthdata.org/gbd-results/ (2024)
In 2021, global DALYs totaled approximately 2.88 billion, comprising 1.75 billion Years of Life Lost (YLL) and 1.13 billion Years Lived with Disability (YLD). This represents a 13% increase from 2019 (2.55B DALYs), largely attributable to COVID-19 deaths and aging populations. YLD accounts for approximately 39% of total DALYs, reflecting the substantial burden of non-fatal chronic conditions. Additional sources: https://vizhub.healthdata.org/gbd-results/ | https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)00757-8/fulltext | https://www.healthdata.org/research-analysis/about-gbd

36. Costs of War Project, Brown University Watson Institute. Environmental cost of war ($100B annually). *Brown Watson Costs of War: Environmental Cost* https://watson.brown.edu/costsofwar/costs/social/environment
War on Terror emissions: 1.2B metric tons GHG (equivalent to 257M cars/year) Military: 5.5% of global GHG emissions (2X aviation + shipping combined) US DoD: World's single largest institutional oil consumer, 47th largest emitter if nation Cleanup costs: $500B+ for military contaminated sites Gaza war environmental damage: $56.4B; landmine clearance: $34.6B expected Climate finance gap: Rich nations spend 30X more on military than climate finance Note: Military activities cause massive environmental damage through GHG emissions, toxic contamination, and long-term cleanup costs far exceeding current climate finance commitments Additional sources: https://watson.brown.edu/costsofwar/costs/social/environment | https://earth.org/environmental-costs-of-wars/ | https://transformdefence.org/transformdefence/stats/

37. ScienceDaily. Medical research lives saved annually (4.2 million). *ScienceDaily: Physical Activity Prevents 4M Deaths* https://www.sciencedaily.com/releases/2020/06/200617194510.htm (2020)
Physical activity: 3.9M early deaths averted annually worldwide (15% lower premature deaths than without) COVID vaccines (2020-2024): 2.533M deaths averted, 14.8M life-years preserved; first year alone: 14.4M deaths prevented Cardiovascular prevention: 3 interventions could delay 94.3M deaths over 25 years (antihypertensives alone: 39.4M) Pandemic research response: Millions of deaths averted through rapid vaccine/drug development Additional sources: https://www.sciencedaily.com/releases/2020/06/200617194510.htm | https://pmc.ncbi.nlm.nih.gov/articles/PMC9537923/ | https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.118.038160 | https://pmc.ncbi.nlm.nih.gov/articles/PMC9464102/

38. SIPRI. 36:1 disparity ratio of spending on weapons over cures. *SIPRI: Military Spending* https://www.sipri.org/commentary/blog/2016/opportunity-cost-world-military-spending (2016)
Global military spending: $2.7 trillion (2024, SIPRI) Global government medical research: $68 billion (2024) Actual ratio: 39.7:1 in favor of weapons over medical research Military R&D alone: $85B (2004 data, 10% of global R&D) Military spending increases crowd out health: 1% ↑ military = 0.62% ↓ health spending Note: Ratio actually worse than 36:1. Each 1% increase in military spending reduces health spending by 0.62%, with effect more intense in poorer countries (0.962% reduction) Additional sources: https://www.sipri.org/commentary/blog/2016/opportunity-cost-world-military-spending | https://pmc.ncbi.nlm.nih.gov/articles/PMC9174441/ | https://www.congress.gov/crs-product/R45403

39. Think by Numbers. Lost human capital due to war ($270B annually). *Think by Numbers* https://thinkbynumbers.org/military/war/the-economic-case-for-peace-a-comprehensive-financial-analysis/ (2021)
Lost human capital from war: $300B annually (economic impact of losing skilled/productive individuals to conflict) Broader conflict/violence cost: $14T/year globally 1.4M violent deaths/year; conflict holds back economic development, causes instability, widens inequality, erodes human capital 2002: 48.4M DALYs lost from 1.6M violence deaths = $151B economic value (2000 USD) Economic toll includes: commodity prices, inflation, supply chain disruption, declining output, lost human capital Additional sources: https://thinkbynumbers.org/military/war/the-economic-case-for-peace-a-comprehensive-financial-analysis/ | https://www.weforum.org/stories/2021/02/war-violence-costs-each-human-5-a-day/ | https://pubmed.ncbi.nlm.nih.gov/19115548/

40. PubMed. Psychological impact of war cost ($100B annually). *PubMed: Economic Burden of PTSD* https://pubmed.ncbi.nlm.nih.gov/35485933/
PTSD economic burden (2018 U.S.): $232.2B total ($189.5B civilian, $42.7B military) Civilian costs driven by: Direct healthcare ($66B), unemployment ($42.7B) Military costs driven by: Disability ($17.8B), direct healthcare ($10.1B) Exceeds costs of other mental health conditions (anxiety, depression) War-exposed populations: 2-3X higher rates of anxiety, depression, PTSD; women and children most vulnerable Note: Actual burden $232B, significantly higher than "$100B" claimed Additional sources: https://pubmed.ncbi.nlm.nih.gov/35485933/ | https://news.va.gov/103611/study-national-economic-burden-of-ptsd-staggering/ | https://pmc.ncbi.nlm.nih.gov/articles/PMC9957523/

41. CGDev. UNHCR average refugee support cost. *CGDev* https://www.cgdev.org/blog/costs-hosting-refugees-oecd-countries-and-why-uk-outlier (2024)
The average cost of supporting a refugee is $1,384 per year. This represents total host country. costs (housing, healthcare, education, security). OECD countries average $6,100 per refugee (mean 2022-2023), with developing countries spending $700-1,000. Global weighted average of $1,384 is reasonable given that 75-85% of refugees are in low/middle-income countries. Additional sources: https://www.cgdev.org/blog/costs-hosting-refugees-oecd-countries-and-why-uk-outlier | https://www.unhcr.org/sites/default/files/2024-11/UNHCR-WB-global-cost-of-refugee-inclusion-in-host-country-health-systems.pdf

42. World Bank. World bank trade disruption cost from conflict. *World Bank* https://www.worldbank.org/en/topic/trade/publication/trading-away-from-conflict
Estimated $616B annual cost from conflict-related trade disruption. World Bank research. shows civil war costs an average developing country 30 years of GDP growth, with 20 years needed for trade to return to pre-war levels. Trade disputes analysis shows tariff escalation could reduce global exports by up to $674 billion. Additional sources: https://www.worldbank.org/en/topic/trade/publication/trading-away-from-conflict | https://www.nber.org/papers/w11565 | http://blogs.worldbank.org/en/trade/impacts-global-trade-and-income-current-trade-disputes

43. VA. Veteran healthcare cost projections. *VA* https://department.va.gov/wp-content/uploads/2025/06/2026-Budget-in-Brief.pdf (2026)
VA budget: $441.3B requested for FY 2026 (10% increase). Disability compensation: $165.6B in FY 2024 for 6.7M veterans. PACT Act projected to increase spending by $300B between 2022-2031. Costs under Toxic Exposures Fund: $20B (2024), $30.4B (2025), $52.6B (2026). Additional sources: https://department.va.gov/wp-content/uploads/2025/06/2026-Budget-in-Brief.pdf | https://www.cbo.gov/publication/45615 | https://www.legion.org/information-center/news/veterans-healthcare/2025/june/va-budget-tops-400-billion-for-2025-from-higher-spending-on-mandated-benefits-medical-care

44. IQVIA Institute for Human Data Science. The global use of medicines 2024: Outlook to 2028. *IQVIA Institute Report* https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/the-global-use-of-medicines-2024-outlook-to-2028 (2024)
Global days of therapy reached 1.8 trillion in 2019 (234 defined daily doses per person). Diabetes, respiratory, CVD, and cancer account for 71 percent of medicine use. Projected to reach 3.8 trillion DDDs by 2028.

45. Sinn, M. P. Private industry clinical trial spending estimate. (2025)
Estimated private pharmaceutical and biotech clinical trial spending is approximately $75-90 billion annually, representing roughly 90% of global clinical trial spending.

46. Cybersecurity Ventures. Cybercrime economy projected to reach $10.5 trillion. *Cybersecurity Ventures: $10.5T Cybercrime* https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/ (2016)
Global cybercrime costs: $3T (2015) → $6T (2021) → $10.5T (2025 projected) 15% annual. growth rate If measured as country, would be 3rd largest economy after US and China Greatest transfer of economic wealth in history Note: More profitable than global trade of all major illegal drugs combined. Includes data theft, productivity loss, IP theft, fraud Additional sources: <https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/> | https://www.boisestate.edu/cybersecurity/2022/06/16/cybercrime-to-cost-the-world-10-5-trillion-annually-by-2025/

47. Sinn, M. P. *The Political Dysfunction Tax.* https://manual.warondisease.org/knowledge/appendix/political-dysfunction-tax.html (2025) doi:10.5281/zenodo.18603840
Quantifying the gap between current global governance and theoretical maximum welfare, estimating a 31-53% efficiency score and $97 trillion in annual opportunity costs.

48. Applied Clinical Trials. Global government spending on interventional clinical trials: $3-6 billion/year. *Applied Clinical Trials* https://www.appliedclinicaltrialsonline.com/view/sizing-clinical-research-market
Estimated range based on NIH ($0.8-5.6B), NIHR ($1.6B total budget), and EU funding. ($1.3B/year). Roughly 5-10% of global market. Additional sources: https://www.appliedclinicaltrialsonline.com/view/sizing-clinical-research-market | https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(20)30357-0/fulltext

49. UBS. Credit suisse global wealth report 2023. *Credit Suisse/UBS* https://www.ubs.com/global/en/family-office-uhnw/reports/global-wealth-report-2023.html (2023)
Total global household wealth: USD 454.4 trillion (2022) Wealth declined by USD 11.3 trillion. (-2.4%) in 2022, first decline since 2008 Wealth per adult: USD 84,718 Additional sources: https://www.ubs.com/global/en/family-office-uhnw/reports/global-wealth-report-2023.html

50. Component country budgets. Global government medical research spending ($67.5B, 2023–2024). *See component country budgets: NIH Budget* https://www.nih.gov/about-nih/what-we-do/budget.

51. United Nations Department of Economic and Social Affairs, Population Division. World population prospects 2024: Summary of results. (2024)
The 2024 Revision of the World Population Prospects provides population estimates and projections for 237 countries or areas. Global median age approximately 30.5 years in 2024, reflecting population-weighted average across all regions.

52. SIPRI. Global military spending ($2.72T, 2024). *SIPRI* https://www.sipri.org/publications/2025/sipri-fact-sheets/trends-world-military-expenditure-2024 (2025).

53. Stockholm International Peace Research Institute. Trends in world military expenditure, 2024. (2025).

54. Estimated from major foundation budgets and activities. Nonprofit clinical trial funding estimate.
Nonprofit foundations spend an estimated $2-5 billion annually on clinical trials globally, representing approximately 2-5% of total clinical trial spending.

55. Industry reports: IQVIA. Global pharmaceutical r&d spending.
Total global pharmaceutical R&D spending is approximately $300 billion annually. Clinical trials represent 15-20% of this total ($45-60B), with the remainder going to drug discovery, preclinical research, regulatory affairs, and manufacturing development.

56. UN. Global population reaches 8 billion. *UN: World Population 8 Billion Nov 15 2022* https://www.un.org/en/desa/world-population-reach-8-billion-15-november-2022 (2022)
Milestone: November 15, 2022 (UN World Population Prospects 2022) Day of. Eight Billion" designated by UN Added 1 billion people in just 11 years (2011-2022) Growth rate: Slowest since 1950; fell under 1% in 2020 Future: 15 years to reach 9B (2037); projected peak 10.4B in 2080s Projections: 8.5B (2030), 9.7B (2050), 10.4B (2080-2100 plateau) Note: Milestone reached Nov 2022. Population growth slowing; will take longer to add next billion (15 years vs 11 years) Additional sources: https://www.un.org/en/desa/world-population-reach-8-billion-15-november-2022 | https://www.un.org/en/dayof8billion | https://en.wikipedia.org/wiki/Day_of_Eight_Billion

57. Harvard Kennedy School. 3.5% participation tipping point. *Harvard Kennedy School* https://www.hks.harvard.edu/centers/carr/publications/35-rule-how-small-minority-can-change-world (2020)
The research found that nonviolent campaigns were twice as likely to succeed as violent ones,. and once 3.5% of the population were involved, they were always successful. Chenoweth and Maria Stephan studied the success rates of civil resistance efforts from 1900 to 2006, finding that nonviolent movements attracted, on average, four times as many participants as violent movements and were more likely to succeed. Key finding: Every campaign that mobilized at least 3.5% of the population in sustained protest was successful (in their 1900-2006 dataset) Note: The 3.5% figure is a descriptive statistic from historical analysis, not a guaranteed threshold. One exception (Bahrain 2011-2014 with 6%+ participation) has been identified. The rule applies to regime change, not policy change in democracies. Additional sources: https://www.hks.harvard.edu/centers/carr/publications/35-rule-how-small-minority-can-change-world | https://www.hks.harvard.edu/sites/default/files/2024-05/Erica%20Chenoweth_2020-005.pdf | https://www.bbc.com/future/article/20190513-it-only-takes-35-of-people-to-change-the-world | https://en.wikipedia.org/wiki/3.5%25_rule

58. World Bank. Gross savings (% of GDP). (2024).

59. NHGRI. Human genome project and CRISPR discovery. *NHGRI* https://www.genome.gov/11006929/2003-release-international-consortium-completes-hgp (2003)
Your DNA is 3 billion base pairs Read the entire code (Human Genome Project,. completed 2003) Learned to edit it (CRISPR, discovered 2012) Additional sources: https://www.genome.gov/11006929/2003-release-international-consortium-completes-hgp | https://www.nobelprize.org/prizes/chemistry/2020/press-release/

60. PMC. Only 12% of human interactome targeted. *PMC* https://pmc.ncbi.nlm.nih.gov/articles/PMC10749231/ (2023)
Mapping 350,000+ clinical trials showed that only 12% of the human interactome has ever been. targeted by drugs. Additional sources: https://pmc.ncbi.nlm.nih.gov/articles/PMC10749231/

61. WHO. ICD-10 code count (14,000). *WHO* https://icd.who.int/browse10/2019/en (2019)
The ICD-10 classification contains approximately 14,000 codes for diseases, signs and. symptoms. Additional sources: https://icd.who.int/browse10/2019/en

62. Wikipedia. Longevity escape velocity (LEV) - maximum human life extension potential. *Wikipedia: Longevity Escape Velocity* https://en.wikipedia.org/wiki/Longevity_escape_velocity
Longevity escape velocity: Hypothetical point where medical advances extend life expectancy. faster than time passes Term coined by Aubrey de Grey (biogerontologist) in 2004 paper; concept from David Gobel (Methuselah Foundation) Current progress: Science adds 3 months to lifespan per year; LEV requires adding >1 year per year Sinclair (Harvard): "There is no biological upper limit to age" - first person to live to 150 may already be born De Grey: 50% chance of reaching LEV by mid-to-late 2030s; SENS approach = damage repair rather than slowing damage Kurzweil (2024): LEV by 2029-2035, AI will simulate biological processes to accelerate solutions George Church: LEV "in a decade or two" via age-reversal clinical trials Natural lifespan cap: 120-150 years (Jeanne Calment record: 122); engineering approach could bypass via damage repair Key mechanisms: Epigenetic reprogramming, senolytic drugs, stem cell therapy, gene therapy, AI-driven drug discovery Current record: Jeanne Calment (122 years, 164 days) - record unbroken since 1997 Note: LEV is theoretical but increasingly plausible given demonstrated age reversal in mice (109% lifespan extension) and human cells (30-year epigenetic age reversal) Additional sources: https://en.wikipedia.org/wiki/Longevity_escape_velocity | https://pmc.ncbi.nlm.nih.gov/articles/PMC423155/ | https://www.popularmechanics.com/science/a36712084/can-science-cure-death-longevity/ | https://www.diamandis.com/blog/longevity-escape-velocity

63. OpenSecrets. Lobbyist statistics for washington d.c. *OpenSecrets: Lobbying in US* https://en.wikipedia.org/wiki/Lobbying_in_the_United_States
Registered lobbyists: Over 12,000 (some estimates); 12,281 registered (2013) Former government employees as lobbyists: 2,200+ former federal employees (1998-2004), including 273 former White House staffers, 250 former Congress members & agency heads Congressional revolving door: 43% (86 of 198) lawmakers who left 1998-2004 became lobbyists; currently 59% leaving to private sector work for lobbying/consulting firms/trade groups Executive branch: 8% were registered lobbyists at some point before/after government service Additional sources: https://en.wikipedia.org/wiki/Lobbying_in_the_United_States | https://www.opensecrets.org/revolving-door | https://www.citizen.org/article/revolving-congress/ | https://www.propublica.org/article/we-found-a-staggering-281-lobbyists-whove-worked-in-the-trump-administration

64. MDPI Vaccines. Measles vaccination ROI. *MDPI Vaccines* https://www.mdpi.com/2076-393X/12/11/1210 (2024)
Single measles vaccination: 167:1 benefit-cost ratio. MMR (measles-mumps-rubella) vaccination: 14:1 ROI. Historical US elimination efforts (1966-1974): benefit-cost ratio of 10.3:1 with net benefits exceeding USD 1.1 billion (1972 dollars, or USD 8.0 billion in 2023 dollars). 2-dose MMR programs show direct benefit/cost ratio of 14.2 with net savings of $5.3 billion, and 26.0 from societal perspectives with net savings of $11.6 billion. Additional sources: https://www.mdpi.com/2076-393X/12/11/1210 | https://www.tandfonline.com/doi/full/10.1080/14760584.2024.2367451

65. Gosse, M. E. Assessing cost-effectiveness in healthcare: History of the $50,000 per QALY threshold. *Sustainability Impact Metrics* https://ecocostsvalue.com/EVR/img/references%20others/Gosse%202008%20QALY%20threshold%20financial.pdf (2008).

66. World Health Organization. Mental health global burden. *World Health Organization* https://www.who.int/news/item/28-09-2001-the-world-health-report-2001-mental-disorders-affect-one-in-four-people (2022)
One in four people in the world will be affected by mental or neurological disorders at some point in their lives, representing [approximately] 30% of the global burden of disease. Additional sources: https://www.who.int/news/item/28-09-2001-the-world-health-report-2001-mental-disorders-affect-one-in-four-people

67. Stockholm International Peace Research Institute. Trends in world military expenditure, 2023. (2024).

68. Calculated from Orphanet Journal of Rare Diseases (2024). Diseases getting first effective treatment each year. *Calculated from Orphanet Journal of Rare Diseases (2024)* https://ojrd.biomedcentral.com/articles/10.1186/s13023-024-03398-1 (2024)
Under the current system, approximately 10-15 diseases per year receive their FIRST effective treatment. Calculation: 5% of 7,000 rare diseases (350) have FDA-approved treatment, accumulated over 40 years of the Orphan Drug Act = 9 rare diseases/year. Adding 5-10 non-rare diseases that get first treatments yields 10-20 total. FDA approves 50 drugs/year, but many are for diseases that already have treatments (me-too drugs, second-line therapies). Only 15 represent truly FIRST treatments for previously untreatable conditions.

69. NIH. NIH budget (FY 2025). *NIH* https://www.nih.gov/about-nih/organization/budget (2024)
The budget total of $47.7 billion also includes $1.412 billion derived from PHS Evaluation financing... Additional sources: https://www.nih.gov/about-nih/organization/budget | https://officeofbudget.od.nih.gov/

70. Bentley et al. NIH spending on clinical trials: 3.3%. *Bentley et al.* https://pmc.ncbi.nlm.nih.gov/articles/PMC10349341/ (2023)
NIH spent $8.1 billion on clinical trials for approved drugs (2010-2019), representing 3.3% of relevant NIH spending. Additional sources: https://pmc.ncbi.nlm.nih.gov/articles/PMC10349341/ | https://catalyst.harvard.edu/news/article/nih-spent-8-1b-for-phased-clinical-trials-of-drugs-approved-2010-19-10-of-reported-industry-spending/

71. PMC. Standard medical research ROI ($20k-$100k/QALY). *PMC: Cost-effectiveness Thresholds Used by Study Authors* https://pmc.ncbi.nlm.nih.gov/articles/PMC10114019/ (1990)
Typical cost-effectiveness thresholds for medical interventions in rich countries range from. $50,000 to $150,000 per QALY. The Institute for Clinical and Economic Review (ICER) uses a $100,000-$150,000/QALY threshold for value-based pricing. Between 1990-2021, authors increasingly cited $100,000 (47% by 2020-21) or $150,000 (24% by 2020-21) per QALY as benchmarks for cost-effectiveness. Additional sources: https://pmc.ncbi.nlm.nih.gov/articles/PMC10114019/ | https://icer.org/our-approach/methods-process/cost-effectiveness-the-qaly-and-the-evlyg/

72. Manhattan Institute. RECOVERY trial 82× cost reduction. *Manhattan Institute: Slow Costly Trials* https://manhattan.institute/article/slow-costly-clinical-trials-drag-down-biomedical-breakthroughs
RECOVERY trial: $500 per patient ($20M for 48,000 patients = $417/patient) Typical clinical trial: $41,000 median per-patient cost Cost reduction: 80-82× cheaper ($41,000 ÷ $500 82×) Efficiency: $50 per patient per answer (10 therapeutics tested, 4 effective) Dexamethasone estimated to save >630,000 lives Additional sources: https://manhattan.institute/article/slow-costly-clinical-trials-drag-down-biomedical-breakthroughs | https://pmc.ncbi.nlm.nih.gov/articles/PMC9293394/

73. Trials. Patient willingness to participate in clinical trials. *Trials: Patients' Willingness Survey* https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-015-1105-3
Recent surveys: 49-51% willingness (2020-2022) - dramatic drop from 85% (2019). during COVID-19 pandemic Cancer patients when approached: 88% consented to trials (Royal Marsden Hospital) Study type variation: 44.8% willing for drug trial, 76.2% for diagnostic study Top motivation: "Learning more about my health/medical condition" (67.4%) Top barrier: "Worry about experiencing side effects" (52.6%) Additional sources: https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-015-1105-3 | https://www.appliedclinicaltrialsonline.com/view/industry-forced-to-rethink-patient-participation-in-trials | https://pmc.ncbi.nlm.nih.gov/articles/PMC7183682/

74. Tufts CSDD. Cost of drug development.
Various estimates suggest $1.0 - $2.5 billion to bring a new drug from discovery through FDA approval, spread across 10 years. Tufts Center for the Study of Drug Development often cited for $1.0 - $2.6 billion/drug. Industry reports (IQVIA, Deloitte) also highlight $2+ billion figures.

75. Value in Health. Average lifetime revenue per successful drug. *Value in Health: Sales Revenues for New Therapeutic Agents* https://www.sciencedirect.com/science/article/pii/S1098301524027542
Study of 361 FDA-approved drugs from 1995-2014 (median follow-up 13.2 years): Mean. lifetime revenue: $15.2 billion per drug Median lifetime revenue: $6.7 billion per drug Revenue after 5 years: $3.2 billion (mean) Revenue after 10 years: $9.5 billion (mean) Revenue after 15 years: $19.2 billion (mean) Distribution highly skewed: top 25 drugs (7%) accounted for 38% of total revenue ($2.1T of $5.5T) Additional sources: https://www.sciencedirect.com/science/article/pii/S1098301524027542

76. Lichtenberg, F. R. How many life-years have new drugs saved? A three-way fixed-effects analysis of 66 diseases in 27 countries, 2000-2013. *International Health* **11**, 403–416 (2019)
Using 3-way fixed-effects methodology (disease-country-year) across 66 diseases in 22 countries, this study estimates that drugs launched after 1981 saved 148.7 million life-years in 2013 alone. The regression coefficients for drug launches 0-11 years prior (beta=-0.031, SE=0.008) and 12+ years prior (beta=-0.057, SE=0.013) on years of life lost are highly significant (p<0.0001). Confidence interval for life-years saved: 79.4M-239.8M (95 percent CI) based on propagated standard errors from Table 2.

77. Deloitte. Pharmaceutical r&d return on investment (ROI). *Deloitte: Measuring Pharmaceutical Innovation 2025* https://www.deloitte.com/ch/en/Industries/life-sciences-health-care/research/measuring-return-from-pharmaceutical-innovation.html (2025)
Deloitte's annual study of top 20 pharma companies by R&D spend (2010-2024): 2024. ROI: 5.9% (second year of growth after decade of decline) 2023 ROI: 4.3% (estimated from trend) 2022 ROI: 1.2% (historic low since study began, 13-year low) 2021 ROI: 6.8% (record high, inflated by COVID-19 vaccines/treatments) Long-term trend: Declining for over a decade before 2023 recovery Average R&D cost per asset: $2.3B (2022), $2.23B (2024) These returns (1.2-5.9% range) fall far below typical corporate ROI targets (15-20%) Additional sources: https://www.deloitte.com/ch/en/Industries/life-sciences-health-care/research/measuring-return-from-pharmaceutical-innovation.html | https://www.prnewswire.com/news-releases/deloittes-13th-annual-pharmaceutical-innovation-report-pharma-rd-return-on-investment-falls-in-post-pandemic-market-301738807.html | https://hitconsultant.net/2023/02/16/pharma-rd-roi-falls-to-lowest-level-in-13-years/

78. Nature Reviews Drug Discovery. Drug trial success rate from phase i to approval. *Nature Reviews Drug Discovery: Clinical Success Rates* https://www.nature.com/articles/nrd.2016.136 (2016)
Overall Phase I to approval: 10-12.8% (conventional wisdom 10%, studies show 12.8%). Recent decline: Average LOA now 6.7% for Phase I (2014-2023 data) Leading pharma companies: 14.3% average LOA (range 8-23%) Varies by therapeutic area: Oncology 3.4%, CNS/cardiovascular lowest at Phase III Phase-specific success: Phase I 47-54%, Phase II 28-34%, Phase III 55-70% Note: 12% figure accurate for historical average. Recent data shows decline to 6.7%, with Phase II as primary attrition point (28% success) Additional sources: https://www.nature.com/articles/nrd.2016.136 | https://pmc.ncbi.nlm.nih.gov/articles/PMC6409418/ | https://academic.oup.com/biostatistics/article/20/2/273/4817524

79. SofproMed. Phase 3 cost per trial range. *SofproMed* https://www.sofpromed.com/how-much-does-a-clinical-trial-cost
Phase 3 clinical trials cost between $20 million and $282 million per trial, with significant. variation by therapeutic area and trial complexity. Additional sources: https://www.sofpromed.com/how-much-does-a-clinical-trial-cost | https://www.cbo.gov/publication/57126

80. Ramsberg, J. & Platt, R. Pragmatic trial cost per patient (median $97). *Learning Health Systems* https://pmc.ncbi.nlm.nih.gov/articles/PMC6508852/ (2018)
Meta-analysis of 108 embedded pragmatic clinical trials (2006-2016). The median cost per. patient was $97 (IQR $19–$478), based on 2015 dollars. 25% of trials cost <$19/patient; 10 trials exceeded $1,000/patient. U.S. studies median $187 vs non-U.S. median $27. Additional sources: https://pmc.ncbi.nlm.nih.gov/articles/PMC6508852/

81. WHO. Polio vaccination ROI. *WHO* https://www.who.int/news-room/feature-stories/detail/sustaining-polio-investments-offers-a-high-return (2019)
For every dollar spent, the return on investment is nearly US$ 39." Total investment cost. of US$ 7.5 billion generates projected economic and social benefits of US$ 289.2 billion from sustaining polio assets and integrating them into expanded immunization, surveillance and emergency response programmes across 8 priority countries (Afghanistan, Iraq, Libya, Pakistan, Somalia, Sudan, Syria, Yemen). Additional sources: https://www.who.int/news-room/feature-stories/detail/sustaining-polio-investments-offers-a-high-return

82. ICRC. International campaign to ban landmines (ICBL) - ottawa treaty (1997). *ICRC* https://www.icrc.org/en/doc/resources/documents/article/other/57jpjn.htm (1997)
ICBL: Founded 1992 by 6 NGOs (Handicap International, Human Rights Watch, Medico. International, Mines Advisory Group, Physicians for Human Rights, Vietnam Veterans of America Foundation) Started with ONE staff member: Jody Williams as founding coordinator Grew to 1,000+ organizations in 60 countries by 1997 Ottawa Process: 14 months (October 1996 - December 1997) Convention signed by 122 states on December 3, 1997; entered into force March 1, 1999 Achievement: Nobel Peace Prize 1997 (shared by ICBL and Jody Williams) Government funding context: Canada established $100M CAD Canadian Landmine Fund over 10 years (1997); International donors provided $169M in 1997 for mine action (up from $100M in 1996) Additional sources: https://www.icrc.org/en/doc/resources/documents/article/other/57jpjn.htm | https://en.wikipedia.org/wiki/International_Campaign_to_Ban_Landmines | https://www.nobelprize.org/prizes/peace/1997/summary/ | https://un.org/press/en/1999/19990520.MINES.BRF.html | https://www.the-monitor.org/en-gb/reports/2003/landmine-monitor-2003/mine-action-funding.aspx

83. OpenSecrets. Revolving door: Former members of congress. (2024)
388 former members of Congress are registered as lobbyists. Nearly 5,400 former congressional. staffers have left Capitol Hill to become federal lobbyists in the past 10 years. Additional sources: https://www.opensecrets.org/revolving-door

84. Kinch, M. S. & Griesenauer, R. H. Lost medicines: A longer view of the pharmaceutical industry with the potential to reinvigorate discovery. *Drug Discovery Today* **24**, 875–880 (2019)
Research identified 1,600+ medicines available in 1962. The 1950s represented industry high-water mark with >30 new products in five of ten years; this rate would not be replicated until late 1990s. More than half (880) of these medicines were lost following implementation of Kefauver-Harris Amendment. The peak of 1962 would not be seen again until early 21st century. By 2016 number of organizations actively involved in R&D at level not seen since 1914.

85. Baily, M. N. Pre-1962 drug development costs (baily 1972). *Baily (1972)* https://samizdathealth.org/wp-content/uploads/2020/12/hlthaff.1.2.6.pdf (1972)
Pre-1962: Average cost per new chemical entity (NCE) was $6.5 million (1980 dollars). Inflation-adjusted to 2024 dollars: $6.5M (1980) $22.5M (2024), using CPI multiplier of 3.46× Real cost increase (inflation-adjusted): $22.5M (pre-1962) → $2,600M (2024) = 116× increase Note: This represents the most comprehensive academic estimate of pre-1962 drug development costs based on empirical industry data Additional sources: https://samizdathealth.org/wp-content/uploads/2020/12/hlthaff.1.2.6.pdf

86. Think by Numbers. Pre-1962 physician-led clinical trials. *Think by Numbers: How Many Lives Does FDA Save?* https://thinkbynumbers.org/health/how-many-net-lives-does-the-fda-save/ (1966)
Pre-1962: Physicians could report real-world evidence directly 1962 Drug Amendments replaced "premarket notification" with "premarket approval", requiring extensive efficacy testing Impact: New regulatory clampdown reduced new treatment production by 70%; lifespan growth declined from 4 years/decade to 2 years/decade Drug Efficacy Study Implementation (DESI): NAS/NRC evaluated 3,400+ drugs approved 1938-1962 for safety only; reviewed >3,000 products, >16,000 therapeutic claims FDA has had authority to accept real-world evidence since 1962, clarified by 21st Century Cures Act (2016) Note: Specific "144,000 physicians" figure not verified in sources Additional sources: https://thinkbynumbers.org/health/how-many-net-lives-does-the-fda-save/ | https://www.fda.gov/drugs/enforcement-activities-fda/drug-efficacy-study-implementation-desi | http://www.nasonline.org/about-nas/history/archives/collections/des-1966-1969-1.html

87. GAO. 95% of diseases have 0 FDA-approved treatments. *GAO* https://www.gao.gov/products/gao-25-106774 (2025)
95% of diseases have no treatment Additional sources: https://www.gao.gov/products/gao-25-106774 | https://globalgenes.org/rare-disease-facts/

88. Oren Cass, Manhattan Institute. RECOVERY trial cost per patient. *Oren Cass* https://manhattan.institute/article/slow-costly-clinical-trials-drag-down-biomedical-breakthroughs (2023)
The RECOVERY trial, for example, cost only about $500 per patient... By contrast, the median per-patient cost of a pivotal trial for a new therapeutic is around $41,000. Additional sources: https://manhattan.institute/article/slow-costly-clinical-trials-drag-down-biomedical-breakthroughs

89. NHS England; Águas et al. RECOVERY trial global lives saved (1 million). *NHS England: 1 Million Lives Saved* https://www.england.nhs.uk/2021/03/covid-treatment-developed-in-the-nhs-saves-a-million-lives/ (2021)
Dexamethasone saved 1 million lives worldwide (NHS England estimate, March 2021, 9 months after discovery). UK alone: 22,000 lives saved. Methodology: Águas et al. Nature Communications 2021 estimated 650,000 lives (range: 240,000-1,400,000) for July-December 2020 alone, based on RECOVERY trial mortality reductions (36% for ventilated, 18% for oxygen-only patients) applied to global COVID hospitalizations. June 2020 announcement: Dexamethasone reduced deaths by up to 1/3 (ventilated patients), 1/5 (oxygen patients). Impact immediate: Adopted into standard care globally within hours of announcement. Additional sources: https://www.england.nhs.uk/2021/03/covid-treatment-developed-in-the-nhs-saves-a-million-lives/ | https://www.nature.com/articles/s41467-021-21134-2 | https://pharmaceutical-journal.com/article/news/steroid-has-saved-the-lives-of-one-million-covid-19-patients-worldwide-figures-show | https://www.recoverytrial.net/news/recovery-trial-celebrates-two-year-anniversary-of-life-saving-dexamethasone-result

90. National September 11 Memorial & Museum. September 11 attack facts. (2024)
2,977 people were killed in the September 11, 2001 attacks: 2,753 at the World Trade Center, 184 at the Pentagon, and 40 passengers and crew on United Flight 93 in Shanksville, Pennsylvania.

91. World Bank. World bank singapore economic data. *World Bank* https://data.worldbank.org/country/singapore (2024)
Singapore GDP per capita (2023): $82,000 - among highest in the world Government. spending: 15% of GDP (vs US 38%) Life expectancy: 84.1 years (vs US 77.5 years) Singapore demonstrates that low government spending can coexist with excellent outcomes Additional sources: https://data.worldbank.org/country/singapore

92. International Monetary Fund. IMF singapore government spending data. (2024)
Singapore government spending is approximately 15% of GDP This is 23 percentage points. lower than the United States (38%) Despite lower spending, Singapore achieves excellent outcomes: - Life expectancy: 84.1 years (vs US 77.5) - Low crime, world-class infrastructure, AAA credit rating Additional sources: https://www.imf.org/en/Countries/SGP

93. World Health Organization. WHO life expectancy data by country. (2024)
Life expectancy at birth varies significantly among developed nations: Switzerland: 84.0. years (2023) Singapore: 84.1 years (2023) Japan: 84.3 years (2023) United States: 77.5 years (2023) - 6.5 years below Switzerland, Singapore Global average: 73 years Note: US spends more per capita on healthcare than any other nation, yet achieves lower life expectancy Additional sources: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-life-expectancy-and-healthy-life-expectancy

94. CSIS. Smallpox eradication ROI. *CSIS* https://www.csis.org/analysis/smallpox-eradication-model-global-cooperation.

95. PMC. Contribution of smoking reduction to life expectancy gains. *PMC: Benefits Smoking Cessation Longevity* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447499/ (2012)
Population-level: Up to 14% (9% men, 14% women) of total life expectancy gain since. 1960 due to tobacco control efforts Individual cessation benefits: Quitting at age 35 adds 6.9-8.5 years (men), 6.1-7.7 years (women) vs continuing smokers By cessation age: Age 25-34 = 10 years gained; age 35-44 = 9 years; age 45-54 = 6 years; age 65 = 2.0 years (men), 3.7 years (women) Cessation before age 40: Reduces death risk by 90% Long-term cessation: 10+ years yields survival comparable to never smokers, averts 10 years of life lost Recent cessation: <3 years averts 5 years of life lost Additional sources: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447499/ | https://www.cdc.gov/pcd/issues/2012/11_0295.htm | https://www.ajpmonline.org/article/S0749-3797(24)00217-4/fulltext | https://www.nejm.org/doi/full/10.1056/NEJMsa1211128

96. ICER. Value per QALY (standard economic value). *ICER* https://icer.org/wp-content/uploads/2024/02/Reference-Case-4.3.25.pdf (2024)
Standard economic value per QALY: $100,000–$150,000. This is the US and global standard. willingness-to-pay threshold for interventions that add costs. Dominant interventions (those that save money while improving health) are favorable regardless of this threshold. Additional sources: https://icer.org/wp-content/uploads/2024/02/Reference-Case-4.3.25.pdf

97. GAO. Annual cost of u.s. Sugar subsidies. *GAO: Sugar Program* https://www.gao.gov/products/gao-24-106144
Consumer costs: $2.5-3.5 billion per year (GAO estimate) Net economic cost: $1. billion per year 2022: US consumers paid 2X world price for sugar Program costs $3-4 billion/year but no federal budget impact (costs passed directly to consumers via higher prices) Employment impact: 10,000-20,000 manufacturing jobs lost annually in sugar-reliant industries (confectionery, etc.) Multiple studies confirm: Sweetener Users Association ($2.9-3.5B), AEI ($2.4B consumer cost), Beghin & Elobeid ($2.9-3.5B consumer surplus) Additional sources: https://www.gao.gov/products/gao-24-106144 | https://www.heritage.org/agriculture/report/the-us-sugar-program-bad-consumers-bad-agriculture-and-bad-america | https://www.aei.org/articles/the-u-s-spends-4-billion-a-year-subsidizing-stalinist-style-domestic-sugar-production/

98. World Bank. Swiss military budget as percentage of GDP. *World Bank: Military Expenditure* https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS?locations=CH
2023: 0.70272% of GDP (World Bank) 2024: CHF 5.95 billion official military spending. When including militia system costs: 1% GDP (CHF 8.75B) Comparison: Near bottom in Europe; only Ireland, Malta, Moldova spend less (excluding microstates with no armies) Additional sources: https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS?locations=CH | https://www.avenir-suisse.ch/en/blog-defence-spending-switzerland-is-in-better-shape-than-it-seems/ | https://tradingeconomics.com/switzerland/military-expenditure-percent-of-gdp-wb-data.html

99. World Bank. Switzerland vs. US GDP per capita comparison. *World Bank: Switzerland GDP Per Capita* https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=CH
2024 GDP per capita (PPP-adjusted): Switzerland $93,819 vs United States $75,492 Switzerland's GDP per capita 24% higher than US when adjusted for purchasing power parity Nominal 2024: Switzerland $103,670 vs US $85,810 Additional sources: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=CH | https://tradingeconomics.com/switzerland/gdp-per-capita-ppp | https://www.theglobaleconomy.com/USA/gdp_per_capita_ppp/

100. OECD. 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

101. OECD. OECD median household income comparison. (2024)
Median household disposable income varies significantly across OECD nations: United. States: $77,500 (2023) Switzerland: $55,000 PPP-adjusted (lower nominal but comparable purchasing power) Singapore: $75,000 PPP-adjusted Additional sources: https://data.oecd.org/hha/household-disposable-income.htm

102. Cato Institute. Chance of dying from terrorism statistic. *Cato Institute: Terrorism and Immigration Risk Analysis* https://www.cato.org/policy-analysis/terrorism-immigration-risk-analysis
Chance of American dying in foreign-born terrorist attack: 1 in 3.6 million per year (1975-2015) Including 9/11 deaths; annual murder rate is 253x higher than terrorism death rate More likely to die from lightning strike than foreign terrorism Note: Comprehensive 41-year study shows terrorism risk is extremely low compared to everyday dangers Additional sources: https://www.cato.org/policy-analysis/terrorism-immigration-risk-analysis | https://www.nbcnews.com/news/us-news/you-re-more-likely-die-choking-be-killed-foreign-terrorists-n715141

103. Wikipedia. Thalidomide scandal: Worldwide cases and mortality. *Wikipedia* https://en.wikipedia.org/wiki/Thalidomide_scandal
The total number of embryos affected by the use of thalidomide during pregnancy is estimated at 10,000, of whom about 40% died around the time of birth. More than 10,000 children in 46 countries were born with deformities such as phocomelia. Additional sources: https://en.wikipedia.org/wiki/Thalidomide_scandal

104. PLOS One. Health and quality of life of thalidomide survivors as they age. *PLOS One* https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0210222 (2019)
Study of thalidomide survivors documenting ongoing disability impacts, quality of life, and long-term health outcomes. Survivors (now in their 60s) continue to experience significant disability from limb deformities, organ damage, and other effects. Additional sources: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0210222

105. US Census Bureau. Historical world population estimates. *US Census Bureau* https://www.census.gov/data/tables/time-series/demo/international-programs/historical-est-worldpop.html
US Census Bureau historical estimates of world population by country and region. (1950-2050). US population in 1960: 180 million of 3 billion worldwide (6%). Additional sources: https://www.census.gov/data/tables/time-series/demo/international-programs/historical-est-worldpop.html

106. FDA Study via NCBI. Trial costs, FDA study. *FDA Study via NCBI* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6248200/
Overall, the 138 clinical trials had an estimated median (IQR) cost of $19.0 million ($12.2. million-$33.1 million)... The clinical trials cost a median (IQR) of $41,117 ($31,802-$82,362) per patient. Additional sources: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6248200/

107. GBD 2019 Diseases and Injuries Collaborators. Global burden of disease study 2019: Disability weights. *The Lancet* **396**, 1204–1222 (2020)
Disability weights for 235 health states used in Global Burden of Disease calculations. Weights range from 0 (perfect health) to 1 (death equivalent). Chronic conditions like diabetes (0.05-0.35), COPD (0.04-0.41), depression (0.15-0.66), and cardiovascular disease (0.04-0.57) show substantial variation by severity. Treatment typically reduces disability weights by 50-80 percent for manageable chronic conditions.

108. WHO. Annual global economic burden of alzheimer's and other dementias. *WHO: Dementia Fact Sheet* https://www.who.int/news-room/fact-sheets/detail/dementia (2019)
Global cost: $1.3 trillion (2019 WHO-commissioned study) 50% from informal caregivers. (family/friends, 5 hrs/day) 74% of costs in high-income countries despite 61% of patients in LMICs $818B (2010) → $1T (2018) → $1.3T (2019) - rapid growth Note: Costs increased 35% from 2010-2015 alone. Informal care represents massive hidden economic burden Additional sources: https://www.who.int/news-room/fact-sheets/detail/dementia | https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12901

109. JAMA Oncology. Annual global economic burden of cancer. *JAMA Oncology: Global Cost 2020-2050* https://jamanetwork.com/journals/jamaoncology/fullarticle/2801798 (2020)
2020-2050 projection: $25.2 trillion total ($840B/year average) 2010 annual cost: $1.16. trillion (direct costs only) Recent estimate: $3 trillion/year (all costs included) Top 5 cancers: lung (15.4%), colon/rectum (10.9%), breast (7.7%), liver (6.5%), leukemia (6.3%) Note: China/US account for 45% of global burden; 75% of deaths in LMICs but only 50.0% of economic cost Additional sources: https://jamanetwork.com/journals/jamaoncology/fullarticle/2801798 | https://www.nature.com/articles/d41586-023-00634-9

110. CDC. U.s. Chronic disease healthcare spending. *CDC* https://www.cdc.gov/chronicdisease/data-research/facts-stats/index.html
Chronic diseases account for 90% of U.S. healthcare spending ($3.7T/year). Additional. sources: https://www.cdc.gov/chronic-disease/data-research/facts-stats/index.html

111. Diabetes Care. Annual global economic burden of diabetes. *Diabetes Care: Global Economic Burden* https://diabetesjournals.org/care/article/41/5/963/36522/Global-Economic-Burden-of-Diabetes-in-Adults
2015: $1.3 trillion (1.8% of global GDP) 2030 projections: $2.1T-2.5T de-. pending on scenario IDF health expenditure: $760B (2019) → $845B (2045 projected) 2/3 direct medical costs ($857B), 1/3 indirect costs (lost productivity) Note: Costs growing rapidly; expected to exceed $2T by 2030 Additional sources: https://diabetesjournals.org/care/article/41/5/963/36522/Global-Economic-Burden-of-Diabetes-in-Adults | https://doi.org/10.1016/S2213-8587(17)30097-9

112. CBO. *The 2024 Long-Term Budget Outlook*. https://www.cbo.gov/publication/60039 (2024).

113. World Bank, Bureau of Economic Analysis. US GDP 2024 ($28.78 trillion). *World Bank* https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US (2024)
US GDP reached $28.78 trillion in 2024, representing approximately 26% of global. GDP. Additional sources: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US | https://www.bea.gov/news/2024/gross-domestic-product-fourth-quarter-and-year-2024-advance-estimate

114. Environmental Working Group. US farm subsidy database and analysis. *Environmental Working Group* https://farm.ewg.org/ (2024)
US agricultural subsidies total approximately $30 billion annually, but create much larger. economic distortions. Top 10% of farms receive 78% of subsidies, benefits concentrated in commodity crops (corn, soy, wheat, cotton), environmental damage from monoculture incentivized, and overall deadweight loss estimated at $50-120 billion annually. Additional sources: https://farm.ewg.org/ | https://www.ers.usda.gov/topics/farm-economy/farm-sector-income-finances/government-payments-the-safety-net/

115. Drug Policy Alliance. The drug war by the numbers. (2021)
Since 1971, the war on drugs has cost the United States an estimated $1 trillion in enforcement. The federal drug control budget was $41 billion in 2022. Mass incarceration costs the U.S. at least $182 billion every year, with over $450 billion spent to incarcerate individuals on drug charges in federal prisons.

116. International Monetary Fund. IMF fossil fuel subsidies data: 2023 update. (2023)
Globally, fossil fuel subsidies were $7 trillion in 2022 or 7.1 percent of GDP. The United States subsidies totaled $649 billion. Underpricing for local air pollution costs and climate damages are the largest contributor, accounting for about 30 percent each.

117. Papanicolas, Irene 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

118. Hsieh, C.-T. & Moretti, E. Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics* https://www.aeaweb.org/articles?id=10.1257/mac.20170388 (2019)
We quantify the amount of spatial misallocation of labor across US cities and its aggregate. costs. Tight land-use restrictions in high-productivity cities like New York, San Francisco, and Boston lowered aggregate US growth by 36% from 1964 to 2009. Local constraints on housing supply have had enormous effects on the national economy. Additional sources: https://www.aeaweb.org/articles?id=10.1257/mac.20170388

119. Yale Budget Lab. The fiscal, economic, and distributional effects of all u.s. tariffs. (2025)
Accounting for all the 2025 US tariffs and retaliation implemented to date, the level of real GDP is persistently -0.6% smaller in the long run, the equivalent of $160 billion 2024$ annually.

120. Tax Foundation. Tax compliance costs the US economy $546 billion annually. https://taxfoundation.org/data/all/federal/irs-tax-compliance-costs/ (2024)
Americans will spend over 7.9 billion hours complying with IRS tax filing and reporting requirements in 2024. This costs the economy roughly $413 billion in lost productivity. In addition, the IRS estimates that Americans spend roughly $133 billion annually in out-of-pocket costs, bringing the total compliance costs to $546 billion, or nearly 2 percent of GDP.

121. Cook, C., Cole, G., Asaria, P., Jabbour, R. & Francis, D. P. Annual global economic burden of heart disease. *International Journal of Cardiology* https://www.internationaljournalofcardiology.com/article/S0167-5273(13)02238-9/abstract (2014)
Heart failure alone: $108 billion/year (2012 global analysis, 197 countries) US CVD: $555B. (2016) → projected $1.8T by 2050 LMICs total CVD loss: $3.7T cumulative (2011-2015, 5-year period) CVD is costliest disease category in most developed nations Note: No single $2.1T global figure found; estimates vary widely by scope and year Additional sources: https://www.ahajournals.org/doi/10.1161/CIR.0000000000001258

122. Source: US Life Expectancy FDA Budget 1543-2019 CSV. US life expectancy growth 1880-1960: 3.82 years per decade. (2019)
Pre-1962: 3.82 years/decade Post-1962: 1.54 years/decade Reduction: 60% decline in. life expectancy growth rate Additional sources: https://ourworldindata.org/life-expectancy | https://www.mortality.org/ | https://www.cdc.gov/nchs/nvss/mortality_tables.htm

123. Source: US Life Expectancy FDA Budget 1543-2019 CSV. Post-1962 slowdown in life expectancy gains. (2019)
Pre-1962 (1880-1960): 3.82 years/decade Post-1962 (1962-2019): 1.54 years/decade Reduction: 60% decline Temporal correlation: Slowdown occurred immediately after 1962 Kefauver-Harris Amendment Additional sources: https://ourworldindata.org/life-expectancy | https://www.mortality.org/ | https://www.cdc.gov/nchs/nvss/mortality_tables.htm

124. Centers for Disease Control and Prevention. US life expectancy 2023. (2024)
US life expectancy at birth was 77.5 years in 2023 Male life expectancy: 74.8 years Female. life expectancy: 80.2 years This is 6-7 years lower than peer developed nations despite higher healthcare spending Additional sources: https://www.cdc.gov/nchs/fastats/life-expectancy.htm

125. US Census Bureau. US median household income 2023. (2024)
US median household income was $77,500 in 2023 Real median household income declined 0.8% from 2022 Gini index: 0.467 (income inequality measure) Additional sources: https://www.census.gov/library/publications/2024/demo/p60-282.html

126. Manuel, D. U.s. Defense spending history: 100 years of military budgets. *DaveManuel.com* https://www.davemanuel.com/us-defense-spending-history-military-budget-data.php (2025)
US military spending in constant 2024 dollars: 1939 $29B (pre-WW2 baseline), 1940 $37B, 1944 $1,383B, 1945 $1,420B (peak), 1946 $674B, 1947 $176B, 1948 $117B, 2024 $886B. The post-WW2 demobilization cut spending 88% in two years (1945-1947). Current peacetime spending ($886B) is 30x the pre-WW2 baseline and 62% of peak WW2 spending, in inflation-adjusted dollars.

127. Statista. US military budget as percentage of GDP. *Statista* https://www.statista.com/statistics/262742/countries-with-the-highest-military-spending/ (2024)
U.S. military spending amounted to 3.5% of GDP in 2024. In 2024, the U.S. spent. nearly $1 trillion on its military budget, equal to 3.4% of GDP. Additional sources: https://www.statista.com/statistics/262742/countries-with-the-highest-military-spending/ | https://www.sipri.org/sites/default/files/2025-04/2504_fs_milex_2024.pdf

128. US Census Bureau. Number of registered or eligible voters in the u.s. *US Census Bureau* https://www.census.gov/newsroom/press-releases/2025/2024-presidential-election-voting-registration-tables.html (2024)
73.6% (or 174 million people) of the citizen voting-age population was registered to vote. in 2024 (Census Bureau). More than 211 million citizens were active registered voters (86.6% of citizen voting age population) according to the Election Assistance Commission. Additional sources: https://www.census.gov/newsroom/press-releases/2025/2024-presidential-election-voting-registration-tables.html | https://www.eac.gov/news/2025/06/30/us-election-assistance-commission-releases-2024-election-administration-and-voting

129. U.S. Senate. Treaties. *U.S. Senate* https://www.senate.gov/about/powers-procedures/treaties.htm
The Constitution provides that the president 'shall have Power, by and with the Advice. and Consent of the Senate, to make Treaties, provided two-thirds of the Senators present concur' (Article II, section 2). Treaties are formal agreements with foreign nations that require two-thirds Senate approval. 67 senators (two-thirds of 100) must vote to ratify a treaty for it to take effect. Additional sources: https://www.senate.gov/about/powers-procedures/treaties.htm

130. Federal Election Commission. Statistical summary of 24-month campaign activity of the 2023-2024 election cycle. (2023)
Presidential candidates raised $2 billion; House and Senate candidates raised $3.8 billion. and spent $3.7 billion; PACs raised $15.7 billion and spent $15.5 billion. Total federal campaign spending approximately $20 billion. Additional sources: https://www.fec.gov/updates/statistical-summary-of-24-month-campaign-activity-of-the-2023-2024-election-cycle/

131. OpenSecrets. Federal lobbying hit record $4.4 billion in 2024. (2024)
Total federal lobbying reached record $4.4 billion in 2024. The $150 million increase in lobbying. continues an upward trend that began in 2016. Additional sources: https://www.opensecrets.org/news/2025/02/federal-lobbying-set-new-record-in-2024/

132. Columbia/NBER. Odds of a single vote being decisive in a u.s. Presidential election. *Columbia/NBER: What Is the Probability Your Vote Will Make a Difference?* https://sites.stat.columbia.edu/gelman/research/published/probdecisive2.pdf (2012)
National average: 1 in 60 million chance (2008 election analysis by Gelman, Silver, Edlin). Swing states (NM, VA, NH, CO): 1 in 10 million chance Non-competitive states: 34 states >1 in 100 million odds; 20 states >1 in 1 billion Washington DC: 1 in 490 billion odds Methodology: Probability state is necessary for electoral college win × probability state vote is tied Additional sources: https://sites.stat.columbia.edu/gelman/research/published/probdecisive2.pdf | https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1465-7295.2010.00272.x

133. Hutchinson and Kirk. Valley of death in drug development. (2011)
The overall failure rate of drugs that passed into Phase 1 trials to final approval is 90%. This lack of translation from promising preclinical findings to success in human trials is known as the "valley of death." Estimated 30-50% of promising compounds never proceed to Phase 2/3 trials primarily due to funding barriers rather than scientific failure. The late-stage attrition rate for oncology drugs is as high as 70% in Phase II and 59% in Phase III trials.

134. DOT. DOT value of statistical life ($13.6M). *DOT: VSL Guidance 2024* https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-on-valuation-of-a-statistical-life-in-economic-analysis (2024)
Current VSL (2024): $13.7 million (updated from $13.6M) Used in cost-benefit analyses for transportation regulations and infrastructure Methodology updated in 2013 guidance, adjusted annually for inflation and real income VSL represents aggregate willingness to pay for safety improvements that reduce fatalities by one Note: DOT has published VSL guidance periodically since 1993. Current $13.7M reflects 2024 inflation/income adjustments Additional sources: https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-on-valuation-of-a-statistical-life-in-economic-analysis | https://www.transportation.gov/regulations/economic-values-used-in-analysis

135. PLOS ONE. Cost per DALY for vitamin a supplementation. *PLOS ONE: Cost-effectiveness of "Golden Mustard" for Treating Vitamin A Deficiency in India (2010)* https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012046 (2010)
India: $23-$50 per DALY averted (least costly intervention, $1,000-$6,100 per death. averted) Sub-Saharan Africa (2022): $220-$860 per DALY (Burkina Faso: $220, Kenya: $550, Nigeria: $860) WHO estimates for Africa: $40 per DALY for fortification, $255 for supplementation Uganda fortification: $18-$82 per DALY (oil: $18, sugar: $82) Note: Wide variation reflects differences in baseline VAD prevalence, coverage levels, and whether intervention is supplementation or fortification Additional sources: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012046 | https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0266495

136. UN News. Clean water & sanitation (LMICs) ROI. *UN News* https://news.un.org/en/story/2014/11/484032 (2014).

137. PMC. Cost-effectiveness threshold ($50,000/QALY). *PMC* https://pmc.ncbi.nlm.nih.gov/articles/PMC5193154/
The $50,000/QALY threshold is widely used in US health economics literature, originating. from dialysis cost benchmarks in the 1980s. In US cost-utility analyses, 77.5% of authors use either $50,000 or $100,000 per QALY as reference points. Most successful health programs cost $3,000-10,000 per QALY. WHO-CHOICE uses GDP per capita multiples (1× GDP/capita = "very cost-effective", 3× GDP/capita = "cost-effective"), which for the US ($70,000 GDP/capita) translates to $70,000-$210,000/QALY thresholds. Additional sources: https://pmc.ncbi.nlm.nih.gov/articles/PMC5193154/ | https://pmc.ncbi.nlm.nih.gov/articles/PMC9278384/

138. Integrated Benefits Institute. Chronic illness workforce productivity loss. *Integrated Benefits Institute 2024* https://www.ibiweb.org/resources/chronic-conditions-in-the-us-workforce-prevalence-trends-and-productivity-impacts (2024)
78.4% of U.S. employees have at least one chronic condition (7% increase since 2021). 58% of employees report physical chronic health conditions 28% of all employees experience productivity loss due to chronic conditions Average productivity loss: $4,798 per employee per year Employees with 3+ chronic conditions miss 7.8 days annually vs 2.2 days for those without Note: 28% productivity loss translates to roughly 11 hours per week (28% of 40-hour workweek) Additional sources: https://www.ibiweb.org/resources/chronic-conditions-in-the-us-workforce-prevalence-trends-and-productivity-impacts | https://www.onemedical.com/mediacenter/study-finds-more-than-half-of-employees-are-living-with-chronic-conditions-including-1-in-3-gen-z-and-millennial-employees/ | https://debeaumont.org/news/2025/poll-the-toll-of-chronic-health-conditions-on-employees-and-workplaces/

139. Duan, N., Kravitz, R. L. & Schmid, C. H. Single-patient (n-of-1) trials: A pragmatic clinical decision methodology. *PubMed* https://pubmed.ncbi.nlm.nih.gov/23849149/ (2013)
Single-patient trials (SPTs, a.k.a. n-of-1 trials) are multiple-period crossover trials conducted within individual patients Application of 2,154 single-patient trials in 108 studies for diverse clinical conditions Conditions addressed: neuropsychiatric (36%), musculoskeletal (21%), pulmonary (13%) Published in Journal of Clinical Epidemiology, 66(8 Suppl): S21-S28 DOI: 10.1016/j.jclinepi.2013.04.006 Additional sources: https://pubmed.ncbi.nlm.nih.gov/23849149/ | https://pmc.ncbi.nlm.nih.gov/articles/PMC3972259/ | https://www.sciencedirect.com/science/article/pii/S089543561300156X

140. Lillie, E. O. *et al.* The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? *PubMed* https://pubmed.ncbi.nlm.nih.gov/21695041/ (2011)
N-of-1 trials consider an individual patient as the sole unit of observation investigating efficacy. or side-effects Goal: determine optimal intervention for individual patient using objective data-driven criteria Can leverage randomization, washout, crossover periods, and placebo controls Argues for serious attention given contemporary focus on individualized medicine Published in Personalized Medicine, 8(2): 161-173. DOI: 10.2217/pme.11.7 Additional sources: https://pubmed.ncbi.nlm.nih.gov/21695041/ | https://pmc.ncbi.nlm.nih.gov/articles/PMC3118090/ | https://www.tandfonline.com/doi/full/10.2217/pme.11.7

141. Ramsberg, J. & Platt, R. Opportunities and barriers for pragmatic embedded trials: Triumphs and tribulations. *Harvard Medical School/Harvard Pilgrim Health Care Institute* https://pmc.ncbi.nlm.nih.gov/articles/PMC6508852/ (2018)
Meta-analysis of 108 embedded pragmatic clinical trials (2006-2016). The median cost per. patient was $97 (mean $478) across all trials reviewed. 25% of studies cost less than $19 per patient. US studies had higher median costs ($187 vs $27 non-US). Registry-based trials were less expensive than EHR-based trials. Traditional RCT comparison: $16,600/patient (Berndt & Cockburn 2014). The 108 trials had median enrollment of 5,540 patients with broad eligibility criteria. 81% used cluster randomization. Trials spanned 15 countries, infectious diseases (25%), cardiovascular (18%), diabetes (12%). Additional sources: https://pmc.ncbi.nlm.nih.gov/articles/PMC6508852/

142. Sakaeda et al. FAERS adverse event underreporting rate. *PubMed: Empirical estimation of under-reporting in FAERS* https://pubmed.ncbi.nlm.nih.gov/28447485/ (2017)
Empirical estimation: Average reporting rate approximately 6%, meaning 94% of adverse. events are underreported Variability: 0.01% to 44% for statin events; 0.002% to >100% for biological drugs; 20% to >100% for narrow therapeutic index (NTI) drugs Selective reporting: Serious, unusual events more likely reported than mild or expected ones Newly marketed drugs: Higher reporting rates due to heightened awareness Older drugs: Events often under-reported Note: FAERS voluntary reporting system captures only "tip of the iceberg" of drug safety problems. Under-reporting introduces inherent biases and limitations in pharmacovigilance data Additional sources: https://pubmed.ncbi.nlm.nih.gov/28447485/ | https://pmc.ncbi.nlm.nih.gov/articles/PMC12393772/

143. Statista / IDC. Wearable device market and adoption statistics. *Statista: Wearable Technology* https://www.statista.com/topics/1556/wearable-technology/ (2024)
Global wearable device users: 1.1 billion in 2024, projected 1.5 billion by 2028. Smartwatch users: 500 million globally (2024) Fitness tracker users: 300 million globally Primary use cases: Health/fitness tracking (sleep, steps, heart rate, activity) Market value: $186 billion (2024), projected $390 billion by 2030 Additional sources: https://www.statista.com/topics/1556/wearable-technology/ | https://www.idc.com/promo/wearablevendor | https://www.insiderintelligence.com/insights/wearable-technology-healthcare-medical-devices/

144. Judea Pearl. *Causality: Models, Reasoning, and Inference*. (Pearl, 2009).
Foundational text on causal inference introducing do-calculus, structural causal models. (SCMs), and graphical causal models Provides mathematical framework for defining and computing causal effects from observational and experimental data Introduces key concepts: interventions (do operator), confounding, counterfactuals, d-separation, causal discovery algorithms First edition 2000, second edition 2009 with new material on counterfactuals and mediation Over 45,000 citations - the seminal work that launched modern causal inference as a discipline Additional sources: https://www.cambridge.org/core/books/causality/B0046844FAE10CBF274D4ACBDAEB5F5B | https://www.amazon.com/Causality-Reasoning-Inference-Judea-Pearl/dp/052189560X | https://scholar.google.com/citations?view_op=view_citation&citation_for_view=bAipNH8AAAAJ:8k81kl-MbHgC

145. Hernán, M. A. & Robins, J. M. *Causal Inference: What If*. (Official Book Website, 2024).
Freely available online textbook on causal inference for scientists who design studies. and analyze data Covers counterfactuals, DAGs, randomized experiments, confounding, selection bias, inverse probability weighting, g-estimation, instrumental variables Publisher: Chapman & Hall/CRC Continuously updated - current version available at author's website Additional sources: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/ | https://content.sph.harvard.edu/wwwhsph/sites/1268/2024/01/hernan-robins_WhatIf_2jan24.pdf | https://remlapmot.github.io/cibookex-r/

146. Austin Bradford Hill. The environment and disease: Association or causation? *PubMed Central: Hill 1965* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/ (1965)
Original paper establishing the 9 criteria for evaluating causal relationships in epidemiology Criteria: Strength, Consistency, Specificity, Temporality, Biological Gradient, Plausibility, Coherence, Experiment, Analogy Published in Proceedings of the Royal Society of Medicine Most influential framework for assessing causation from observational data Additional sources: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/ | https://en.wikipedia.org/wiki/Bradford_Hill_criteria