Epidynamix: From Force to Field in Real-World Epidemiology

Jinseob Kim

Abstract

Standard causal inference methods estimate average treatment effects (ATE) under positivity and no unmeasured confounding assumptions. In real-world data (RWD), structural positivity violations arise from clinical guidelines and contraindications, while treatment effects are often heterogeneous across patient states. We propose Epidynamix, a framework that models health states as points within a Risk Potential Field $\Phi(s) = -\log \lambda(s)$, where treatment effects become directional gradients and positivity violations define structural boundaries. Through simulation, we show that while ATE-based methods (Cox regression, IPTW) collapse effect heterogeneity into scalar summaries, the field-based approach preserves clinically meaningful variation (~4-fold across state space). This geometric perspective offers a complementary lens for RWD analysis, particularly when the standard causal question is structurally unanswerable.

Introduction

For decades, causal inference in epidemiology has operated under a "Newtonian" abstraction: treatment A exerts a force on outcome Y, producing an Average Treatment Effect (ATE) that we estimate as if it were a universal constant.

This framework assumes: (1) interventions are external forces applied to passive subjects; (2) positivity violations are statistical nuisances to be trimmed; and (3) a single scalar (ATE or Hazard Ratio) captures the truth of treatment impact.

Why now? The explosion of Real-World Data (RWD)—electronic health records, claims databases, registries—has exposed the limitations of this paradigm. Unlike curated trial populations, RWD reflects the full complexity of clinical practice: treatment decisions are constrained by guidelines, contraindications create structural zeros in the propensity score, and effects vary systematically across patient states. We increasingly face situations where:

- Treatment is mandated for high-risk patients (structural positivity violation)
- The "same" intervention varies by context, timing, and patient state
- Treatment decisions are entangled with prognosis (confounding by indication)

Traditional methods respond by restricting analysis to "overlap regions" or trimming extreme

weights—effectively discarding the very patients for whom treatment decisions matter most (Petersen et al. 2012).

We propose an alternative: **a map, not a number.**

The Geometric Turn: From Force to Field

We propose a fundamental shift: **Causal effects are not primitive forces; they are movements within a structured Field.** We call this framework Epidynamix.

The Risk Potential Field

We define a scalar field over the state space $\mathcal{S}$ of all possible patient conditions, called the Risk Potential:

$$\Phi(s) = -\log(\lambda(s))$$

where $\lambda(s)$ is the instantaneous hazard at state s. In this field, "safety" is high ground (peaks), and "danger" is low ground (valleys).

Intervention as State-Transition

An intervention is not a force applied to a static object. It is an operator $\mathcal{I}_a$ that moves a patient from one coordinate in state space to another. For binary treatment, the effect is:

$$\delta_A \Phi(x) = \Phi(x, A = 1) - \Phi(x, A = 0)$$

The ATE is simply the population average of this local effect:

$$\text{ATE} = \mathbb{E}_X[\delta_A \Phi(X)]$$

This average is information-preserving only when $\text{Var}_X[\delta_A \Phi] \approx 0$. When effects vary across patient states, the ATE collapses a rich distribution into a single number.

Positivity Violations as Structural Boundaries

In standard causal inference, if a subgroup always receives treatment ($P(A = 1|X) = 1$), we call it a positivity violation—a statistical problem to fix.

In the Epidynamix framework, this is a **Structural Cliff**: a boundary where the untreated state is forbidden by ethics, guidelines, or biological reality. Rather than treating these as estimation failures, we map them as the defining edges of the clinical landscape.

Simulation Study

To demonstrate the Field approach, we simulated a clinical scenario with a structural boundary and heterogeneous effects.

Setup:

- X_1: Systolic blood pressure (100–200 mmHg)
- X_2: Inflammatory marker (0–10 mg/L)
- A: Antihypertensive treatment (binary)
- Structural constraint: If $X_1 > 160$, then $A = 1$ (mandatory per guideline)

Treatment benefit was designed to increase with X_1.

Results (N = 3,000):

Method	Output	Interpretation
Cox HR	0.55	Treatment reduces hazard by 45%
IPTW ATE	0.20	Treatment increases 1-year survival by 20 percentage points
Field $\delta_A\Phi$	0.42–1.66	Effect varies 4-fold across state space

Traditional methods produce single numbers. The Field approach reveals that treatment benefit varies ~4-fold across the state space, with the largest effects in high-BP patients—precisely those for whom the guideline mandates treatment (Cole and Hernán 2008). The region $X_1 > 160$ is not a failure; it is a Structural Cliff revealing clinical logic.

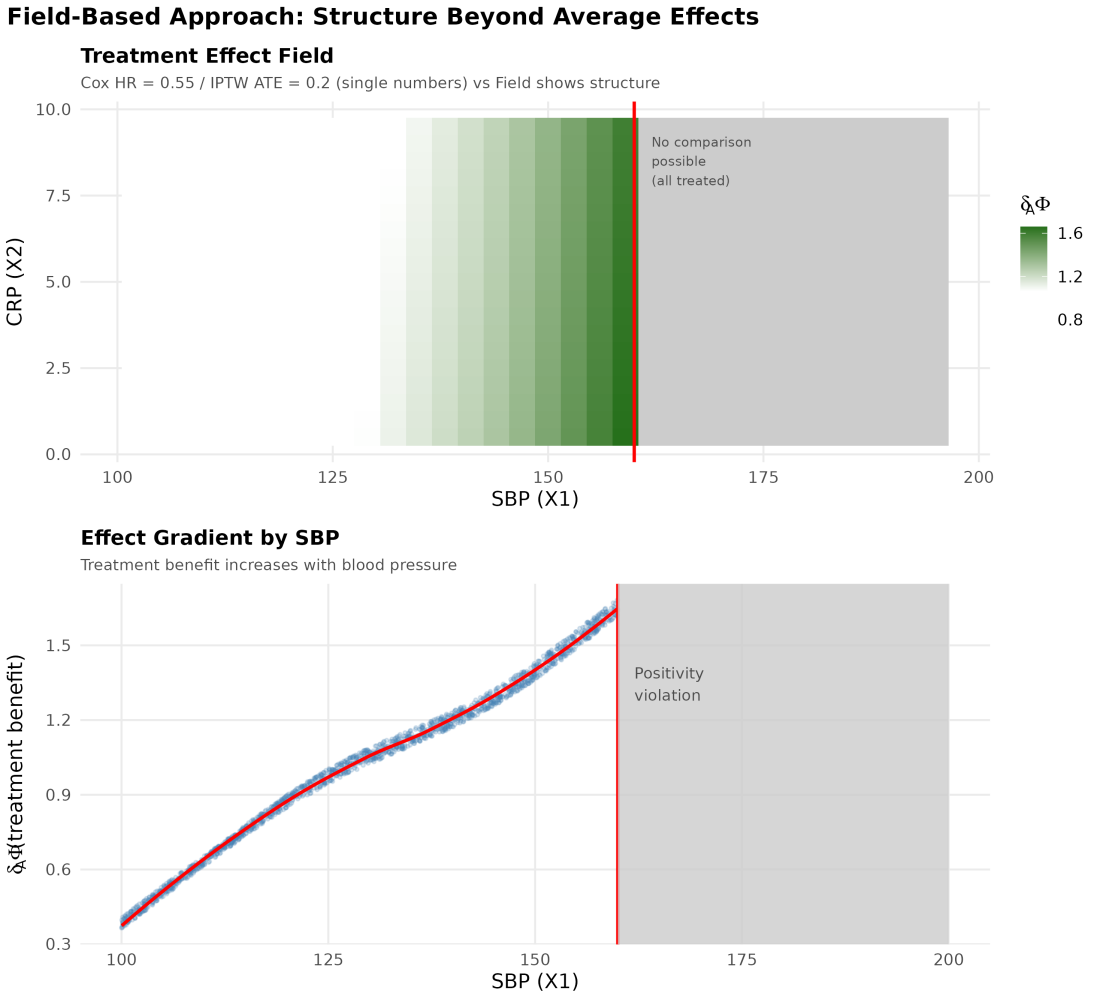

Figure 1: Treatment effect field showing heterogeneity and structural boundary at $X_1 > 160$. Gray region indicates positivity violation where comparison is structurally impossible.

Relation to Existing Approaches

Differentiation from CATE

Conditional Average Treatment Effect (CATE) estimation—via causal forests (Wager and Athey 2018), meta-learners, or Bayesian methods—also addresses heterogeneity. The difference lies in the question asked:

Aspect	CATE	Epidynamix
Question	What is the effect for subgroup $X = x$?	What is the structure of the risk landscape?
Positivity	Estimation problem	Structural information
Counterfactuals	Required	Not required in boundary regions
Output	$\tau(x)$	$\Phi(s)$, gradients, boundaries

CATE asks "what is the effect?"—conditional on covariates. Epidynamix asks "what does the landscape look like, and where can we not go?" Crucially, CATE treats positivity violations as missing data; Epidynamix reports boundaries explicitly as part of the result.

Connection to Structural Nested Models

The local effect $\delta_A \Phi(x)$ shares conceptual ground with the "blip function" in Structural Nested Mean Models (SNMM) (Robins 1994), which also models treatment effects as functions of patient state. The key difference is interpretive: SNMM aims to estimate causal effects under sequential exchangeability, while Epidynamix treats the effect surface as a geometric object to be mapped, with boundaries as informative features rather than obstacles.

Novelty Claim

We do not claim to invent new mathematics. Potential landscapes exist in physics and systems biology (e.g., Waddington's epigenetic landscape (Waddington 1957)). Our contribution is applying this lens to clinical RWD, where positivity violations are common, effects are entangled with state, and the average-effect question may be structurally unanswerable.

This is a reinterpretation, not an invention—a new language for an old problem.

Discussion

The Epidynamix framework offers a complementary perspective to standard causal inference (Rubin 1974; Hernán and Robins 2020). It does not replace existing methods; rather, it clarifies their domain of validity.

When ATE works: If the risk field is smooth, low-curvature, and well-connected (positivity holds everywhere), the ATE is an accurate summary. Traditional methods excel here.

When ATE fails: If effects are highly heterogeneous, or if structural constraints create disconnected regions (Petersen et al. 2012), the ATE may be misleading or undefined. The field approach provides richer output—a map rather than a number.

Practical implications:

1. Report effect heterogeneity alongside average effects
2. Visualize the treatment effect landscape when possible
3. Treat positivity violations as findings, not errors
4. Consider whether the causal question is answerable before estimating

Limitations: The framework is conceptual; practical estimation of $\Phi(s)$ requires flexible models and careful validation. Extension to time-varying treatments and high-dimensional states remains an open challenge.

Conclusion

Newtonian mechanics is not wrong—it is flat-space physics. Causal inference is not false—it is a low-curvature approximation.

Real-world epidemiology, with its structural constraints and heterogeneous effects, may require a geometric view. The Epidynamix framework offers one such lens: treating causal effects not as forces, but as directional movements within a structured risk field.

When the standard causal question cannot be answered, perhaps we should ask a different question.

References

Cole, Stephen R, and Miguel A Hernán. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology* 168 (6): 656–64.

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Petersen, Maya L, Kristin E Porter, Susan Comfort, Rachel Morello-Frosch, and Rose Ray. 2012. "Diagnosing and Responding to Violations in the Positivity Assumption." *Statistical Methods in Medical Research* 21 (1): 31–54.

Robins, James M. 1994. "Correcting for Non-Compliance in Randomized Trials Using Structural Nested Mean Models." *Communications in Statistics-Theory and Methods* 23 (8): 2379–2412.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688.

Waddington, Conrad Hal. 1957. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. London: George Allen & Unwin.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Causal Forests." *Journal of the American Statistical Association* 113 (523): 1228–42.