

ARTICLE

THE MIRAGE OF CORRELATION

www.rgnmed.com | circles@rgnmed.com

THE PREMISE

Modern biomedicine is awash in correlations. High-throughput assays, EHR exhaust, and observational registries generate torrents of associations linking exposures to outcomes. These signals are cheap to produce and easy to publish. Yet clinical action requires something harder: causal understanding. When we mistake correlation for cause, we translate statistical noise into medical advice, inflate hope, and waste scarce clinical attention. The problem is not that correlations are useless; it is that they are increasingly treated as sufficient.

THE DISTORTION

Correlation masquerades as causation through three recurrent pathways:

- **Confounding and selection.** Patients self-select into treatments; clinicians allocate therapies based on prognosis; healthier people seek screening. Unmeasured factors drive both exposure and outcome, creating spurious links. Without a causal design – clear counterfactuals, exchangeability, and temporality – associations reflect the clinic's sorting mechanism more than biology.
- **Surrogate bias.** We optimize for variables that are measurable (biomarkers, intermediate endpoints) rather than variables that matter (morbidity, mortality, function). Surrogates correlate with outcomes in one context and fail in another, inviting ineffective or harmful interventions that “improve the number” while leaving patients unchanged.
- **Flexible analysis and garden-of-forking-paths.** When thousands of features meet dozens of modeling choices, some association will appear significant. Absent prespecified analyses, causal graphs, and sensitivity checks, correlation is a by-product of researcher degrees of freedom, not the world.

Together, these distortions reward speed over design. They generate publishable patterns that crumble at the bedside because they never answered a causal question in the first place.

THE CONSEQUENCE

The mirage of correlation has practical and moral costs:

- **Therapeutic misdirection.** Interventions built on non-causal signals underperform in trials or succeed on surrogates while failing on outcomes, exposing patients to cost and risk without benefit.
- **Policy volatility.** Public guidance oscillates with each new association study, eroding trust among clinicians and citizens who experience “whiplash science.”
- **Equity harms.** Spurious correlations often encode structural confounding (access, environment, bias). Acting on them can amplify disparities by directing resources toward populations easiest to measure rather than those most likely to benefit.
- **Epistemic stagnation.** When correlations are treated as answers, we stop asking mechanistic questions. Biology becomes a backdrop for analytics rather than the governor of inference.

In short, correlation without design produces abundant *signals* but little *understanding* – a surplus of claims and a deficit of care.

THE WAY FORWARD

Restoring causation requires redesign, not rhetoric:

- **Start with a target trial.** In observational settings, explicitly specify the randomized trial you wish you could run: eligibility, treatment strategies, time zero, assignment, outcomes, follow-up, and causal contrasts. Then emulate it with appropriate data and methods.
- **Draw the causal graph.** Make assumptions visible with directed acyclic graphs (DAGs). Identify confounders to adjust, colliders to avoid, mediators to preserve, and instruments to employ. Method follows model.
- **Commit to temporality and specification.** Define exposures and outcomes prospectively where possible; preregister analytic plans; limit researcher degrees of freedom; conduct sensitivity analyses (negative controls, E-values, falsification endpoints).

- **Retire weak surrogates.** Tie intermediate markers to outcomes through validated causal pathways – or prioritize trials and longitudinal endpoints that capture what patients value.
- **Report counterfactuals, not just correlations.** Express results as effects under interventions (“If assigned strategy A vs B...”) and disclose the assumptions under which these effects are identified.
- **Integrate mechanism.** Let biology constrain models. Use causal reasoning to decide *what must be true* for an effect to exist, and design studies that could prove it false.

Correlation is a starting line, not a finish. Medicine earns trust when we move from signals to causes – because only causes tell us what to do.

SELECTED REFERENCES

- RegenMed (2025). *Genuine Medical Research Has Lost Its Way*. White Paper, November 2025.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
- Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.
- Hill, A. B. (1965). *The Environment and Disease: Association or Causation?* Proceedings of the Royal Society of Medicine, 58, 295–300.
- Rubin, D. B. (1974). *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*. Journal of Educational Psychology, 66(5), 688–701.
- Hernán, M. A., & Robins, J. M. (2016). *Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available*. Annals of Internal Medicine, 164(9), 671–677.
- Westreich, D., & Greenland, S. (2013). *The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients*. Epidemiology, 24(2), 282–284.

GET INVOLVED OR LEARN MORE – CONTACT US TODAY!

If you are interested in contributing to this important initiative or learning more about how you can be involved, please contact us*:

 [RegenMed | www.rgnmed.com](http://RegenMed.com)

 circles@rgnmed.com



SCAN ME

*If the links do not work for you, please download the PDF.