Technical Note: The Thin Red Line of Epidemiological Stability. Detecting e-Iatrogenesis in Big Medical Data Big Medical Data

Marco Roccetti

Department of Computer Science and Engineering, University of Bologna, Bologna I-40126, Italy.

Correspondence to:
Marco Roccetti, Department of Computer Science and Engineering, University of Bologna, I-40126, Italy
E-mail: marco.roccetti@unibo.it

Abstract

The rapid transition toward large-scale retrospective Big Data analysis in biomedicine has introduced a critical epistemological paradox: a scenario where statistical precision often acts as a sophisticated mask for a profound lack of clinical accuracy. This Technical Note addresses the rising phenomenon of e-iatrogenesis, defined here as the systematic generation of unintended medical misinformation through the interpretation of unstable bio-data. We argue that sophisticated statistical models, when disconnected from national gold biomedical standards, can inadvertently crystallize epidemiological impossibilities. By examining three exemplar cases, spanning oncology, dermatology, and psychiatry, this work introduces a formal sensitivity stress test, the Sensitivity Hazard Ratio. This bio-computational metric is designed to measure signal instability by benchmarking big data-driven outcomes against national census-level health registries. Our analysis reveals that in all three instances, the reported clinical signals, such as a 77% reduction in schizophrenia or a fourfold increase in vitiligo incidence, were mathematically manufactured by failing denominators or structural selection biases within the data-driven cohorts. When recalibrated against authoritative national benchmarks, these perceived risks effectively dissolve, exposing the thin red line between a genuine epidemiological discovery and a simole computational outcome. We conclude that there is an urgent need for systematic epistemological maintenance of the scientific record. To prevent unstable signals from being fossilized into public health policies and clinical guidelines, we advocate for the mandatory adoption of dynamic bio-informatic audits and rigorous benchmarking as a standard prerequisite for large-scale observational research.

Keywords: Health informatics, e-iatrogenesis, data integrity, epidemiological paradox, sensitivity analysis, big bio-medical data

Highlights

- Large-scale retrospective databases may be prone to e-iatrogenesis through the generation of unstable clinical signals.
- Sophisticated statistical models can inadvertently crystallize epidemiological impossibilities when decoupled from bio-clinical reality.
- The Sensitivity Hazard Ratio is introduced as a formal stress-test to anchor big bio-medical data to national gold standards.
- Case studies in oncology and psychiatry demonstrate how denominator depletion leads to significant signal instability.
- We argue for systematic epistemological maintenance to prevent the sedimentation of unstable data into public bio-health policies.

1. The Epistemological Shift

The landscape of modern epidemiology is undergoing a profound structural reconfiguration, shifting from the controlled environment of prospective trials to the vast, often unmapped territories caracterized by retrospective Big Data studies. This transition from classical, hypothesis-driven clinical research to large-scale retrospective observational investigations has introduced a dangerous epistemological paradox: the gradual erosion of causal inference in favor of automated, algorithmic associations. In the current era of Big Bio-Medical Data, we frequently witness a categorical confusion between statistical precision and scientific accuracy. While massive N-sizes naturally produce narrow confidence intervals and seductive p-values, they do not always and inherently guarantee proximity to the biological or epidemiological truth. Dangerously, a sufficiently large dataset can grant an aura of infallibility to signals that could be, instead, just the outcomes of statistical artifacts or structural biases.

This systemic confusion fuels a phenomenon which has been defined as *e-iatrogenesis* [1]. Unlike more traditional iatrogenesis, which involves direct harm from medical intervention, e-iatrogenesis represents a digital-era pathology: the generation of unintended patient harm or the enactment of misguided public health policies through the integration of erroneous, unverified, or unbalanced data into human and digital decision-support systems. Moreover, when these unstable signals are published in high-impact journals, they undergo a process of digital sedimentation. They cease to be viewed as mere observational hypotheses confined to a single experiment and are instead fossilized into generalized medical facts that influence meta-analyses, insurance algorithms, and the training sets of Large Language Models (LLMs). This creates an illusory feedback loop where unstable data justifies unstable policies, and the sheer volume of the data acts as a sophisticated smokescreen, allowing statistical significance to bypass the fundamental check of clinical and logical plausibility [2].

To restore scientific integrity, bio-medical science must not lose contact with data integrity building upon three strategic pillars designed to act as a firewall against e-iatrogenesis:

- **Dynamic STROBE Audits:** The STROBE checklist should evolve from a passive reporting requirement into a rigorous, dynamic verification method. Authors should be held accountable not just for reporting their methodology, but for proving that their specific cohort's internal validity is not an island separated from the external reality [3].

- **Mandatory Benchmarking:** We argue for the institutionalization of a kind of reality check against the unregulated use of big data. Every retrospective study should be required to calibrate its crude incidence rates against national gold standards or established registries. A study that operates in an epidemiological neutral zone, ignoring the baseline burden of disease in the general population, raises serious concerns if it is used as an undiscussed a reliable basis for health policies.
- **The Clinician's Final Say:** We must reclaim the primacy of clinical sensitivity. The dictatorship of the N-size must not be allowed to silence the experienced bio-clinician who observes that a reported signal, such as a four-fold increase in a rare disease or the near-disappearance of a chronic one (see below), defies historical epidemiological constants and bedside reality.

We argue that the several failures observed when using blindly the current retrospective paradigm is the isolation of the study cohort from the real-world environment. In many Big Bio-Data architectures, the control group incidence (or the unexposed) functions as a hidden degree of freedom. Because retrospective cohorts are often filtered through complex inclusion/exclusion criteria or insurance-claim logic, the resulting unexposed group may suffer from an unrecognized healthy-user bias or structural depletion. If the denominator of a computed Hazard Ratio is set for some reason too low or high, the resulting value becomes a mathematical deception, that is a ratio that measures the eccentricity of the cohort rather than the risk of the exposure.

To measure the extent of this possible gap, we propose to exploit a sensitivity analysis where the observed HR is compared against a calibrated Sensitivity Hazard Ratio. To be more explicit, it is well known that in the classical Cox Proportional Hazards model, the HR is essentially the ratio of the hazard rate in the exposed group, $I(\mathrm{E})$, to the hazard rate in the unexposed group, $I(\mathrm{NE})$, based on the following formula

$$\mathrm{HR} = I(\mathrm{E}) / I(\mathrm{NE}). \tag{1}$$

Our proposed sensitivity test involves replacing the internal, cohort-specific $I(\mathrm{NE})$ with an established *National Gold Benchmark* $I(\mathrm{N})$, derived from official health registries or census-level data. This allows us to verify if the clinical signal oscillates around the thin red line of epidemiological plausibility:

$$\text{Sensitivity HR} = I(\mathrm{E}) / I(\mathrm{N}) \tag{2}$$

By shifting the anchor of the calculation from an isolated control group to a verified population baseline, Sensitivity HR provides the ultimate reality check. If the original statistically significant risk vanishes, approaches 1.0, or even inverts when the denominator nears the national average, the original signal should be considered unstable, as highly variable with respect its baseline. In other words, one could have the suspicion that it could be only a byproduct of selection bias or e-iatrogenic noise, rather than a clear witness of a biological association or medical causality.

Ensuring this constant epistemological maintenance is not merely a technical option; it is a professional imperative. We must prevent a future where medical truth is decided by the loudest algorithm rather than the most faithful data. The thin red line represents the boundary between evidence-based medicine and data-driven empirical theories; as guardians of the scientific record, our duty is to ensure that the metrics we measure never become more important than the reality they are intended to represent. In this regard, in the following Subsections we present three running examples and describe

how their baseline sensitivity is measured, in the sense discussed above

1.1. Case I – A Dermatological Impossibility

The first running example amounts to a high-profile case claiming a significant association between COVID-19 vaccination and new-onset vitiligo [4] based on an immense dataset with almost 3,000,000,000 individuals. The authors reported an incidence of 0.67 per 10,000 in the control group versus 2.22 per 10,000 in the vaccinated group within a time framework of three months. While these numbers yielded a statistically significant Hazard Ratio, they become unstable under the weight of a simple reality check.

When benchmarked against the established national baseline of 2.473 per 10,000 [5], the study's data reveals an impossible temporal paradox. The control non-vaccinated group shows a over 70% deficit in cases within the specific 3-month observational window compared to the national reality. This is not a mere sampling variation; it is a depletion of the denominator. Even more striking is the biological projection: if the vaccinated group truly reached an incidence of 2.22 in such a narrow window, the annualized incidence would skyrocket to approximately 8.8 per 10,000 [6, 7].

To accept this signal as valid, one would have to believe that a medical intervention could suddenly quadruple the incidence of a stable, chronic autoimmune condition across an entire nation, a kind of epidemic of vitiligo that has somehow remained invisible to every national registry and clinical practitioners. This case demonstrates the essence of e-iatrogenesis: the statistical model warns a risk because it compares a nearly-empty control group against a group that, while appearing to have a higher risk, still sits below the national average when normalized. The significance is a mathematical outcome, born from a denominator that has lost all contact with the temporal and epidemiological truth of the country.

1.2. Case II – A Psychiatric Incredibility

The second example [8] shifts the focus to mental health but originates from a similar very large dataset (millions of individuals). The study reports two diametrically opposed but equally impossible signals: a 77% reduction in schizophrenia risk (Hazard Ratio 0.23) and relevant increase in Bipolar Disorder (BD) within just 90 days of vaccination. A quantitative audit [9] reveals structural oscillations that invalidate both signals based on two distinct epidemiological impossibilities.

First, the reported 77% reduction in schizophrenia risk within a 90-day window is clinically unprecedented. Schizophrenia is a chronic neurodevelopmental disorder with an insidious onset; the idea that a vaccine could prevent it in such a massive proportion in just three months lacks any known biological mechanism. This represents a miss in external validity (STROBE 21**)**, where the statistical model interprets a structurally depleted control group as a biological protective effect.

Second, the study’s report on Bipolar Disorder is equally problematic. The incidence reported for a single quarter (90 days) surprisingly accounts for 14% of the total annual prevalence of the disease in South Korea. For a stable chronic condition, such a massive concentration of new cases in a single three-month window is an epidemiological impossibility. These signals are not medical discoveries; but clear cases where the statistical machinery has lost all contact with the baseline reality of psychiatric medicine.

1.3. Case III – An Oncological Paradox

The final is another example of how a very large medical dataset [10] can manifest a total disconnection from national gold standards. By extracting the raw data provided, 12,133 cancer cases within a cohort of 2,975,035, we achieve a Crude Incidence Rate (CR) of 40.78 per 10,000 for the entire study population.

When benchmarked against the national cancer registry, which reports a national CR of 52.46 per 10,000 [11] for the 2022, the entire study cohort shows a 22.3% deficit in cancer diagnoses. This gap equates to approximately 50,000 missing cancer cases in a single year compared to the national reality.

The most alarming instability occurs in the unvaccinated control group, which reports a CR of only 33.43 per 10,000. This group is an extreme statistical outlier, under-reporting cancer by 36.3% relative to the national average. By using such a depleted denominator as the baseline for comparison, the study suggests an increased risk in the vaccinated group (CR 42.63), even though the vaccinated group's incidence itself remains 18.7% below the national gold standard. This is hardly interpretable as a clean biological signal of carcinogenesis; it seems more a direct effect of baseline instability [12].

1.4. Computing the Sensitivity

As already anticipated, the instability of all signals described in the three examples above becomes evident if we subject them to a sensitivity reality check. By benchmarking the reported signals against national gold standards, we demonstrate that the perceived clinical associations questionable as a byproduct of unstable signals which do not bear the test of moving the baseline. In particular, to perform the test, instead of relying solely on the example's internal, and often oscillating, control group incidence *I*(NE), we stress-test the original HR by using the national gold benchmark *I*(N), the verified incidence of the condition in the general population, as the denominator (or numerator in the case of schizophrenia). This transition from an internal Hazard Ratio to a Sensitivity Hazard Ratio serves as the ultimate anchor to reality, as illustrated in Table 1 below [13-15].

Table 1. Sensitivity Stress Test: Recalibration.

Disease	Original HR	I(E) - I(NE)	I(N)	Sensitivity Formula	Sensitivity Results
Vitiligo	2.714	2.22 - 0.67	2.47	I(E)/I(N) = 2.22/2.47	0.89
Schizophrenia	0.231	0.51 - 1.98	2.10	I(N)/I(NE) = 2.1/1.98	1.06
All-Cancers	1.27	42.63 - 33.42	55.02	I(E)/I(N) = 39.43/55.02	0.77

The results of this stress test are conclusive and expose a systematic methodological instability always present in the examined big-data driven examples. In the case of Schizophrenia, the sensitivity check reveals a Sensitivity HR of approximately 1 (2.1 / 1.98 = 1.06) between the national benchmark [5] and the study's control group[4]. In this case, in fact, while the denominator is epidemiologically plausible, the reported HR

of 0.231 is a biological inconsistency: it implies that the vaccine erased 76% of expected natural cases in the vaccinated group (0.51 exposed vs. 2.1 non-exposed). Such an effect size suggests a structural exclusion of prevalent cases rather than a protective clinical effect. As soon the numerator is returned to the national reference value, the protective effect disappears.

More regularly, in the cases of Vitiligo and All Cancers, the Stability Hazard Ratio (> 1) reveals that the reported risks were generated by quite anemic control groups that significantly under-reported cases compared to national reality. Once recalibrated against the *I*(N) benchmark, the vitiligo signal decreases to 0.89 and the oncology risk to 0.77 as exemplar effects of our sensitivity analysis.

Here, It is important to reiterate that this sensitivity calculation is not a counterfactual in the strictest sense, nor is it intended to replace the original HR, which must still be computed using sophisticated adjustment techniques, but rather it serves exclusively as an epidemiological stability test to verify whether the observed signal remains anchored to the reality of national gold standards [16-18].

2. Conclusion

The transition from traditional epidemiology to Big Bio-Data analytics must not be a one-way street toward automated misinformation [19, 20]. As we have demonstrated through the cases of dermatology, psychiatry, and oncology, the current reliance on internal cohort validity, without a rigorous external reality check, is generating a digital-era pathology: *e-iatrogenesis*. When statistical significance is decoupled from clinical and population-level plausibility, the resulting signals do not represent scientific discoveries, but rather mathematical artifacts born from failing denominators or structural selection biases.

The introduction of the Sensitivity Hazard Ratio (Sensitivity HR) is a necessary step toward restoring the integrity of the scientific record. By anchoring internal cohort results to verified national gold standards, we provide a transparent audit tool that distinguishes biological causality from statistical noise. This approach moves beyond the passive reporting of the STROBE checklist, demanding a dynamic verification of the study's external validity.

Furthermore, our findings emphasize that the N-size dictatorship must be balanced by clinical intuition and epidemiological constants. A system that allows a 77% reduction in schizophrenia or a 36% under-reporting of cancer to pass as valid signals is a system in need of urgent epistemological maintenance. Publishers and editorial boards must recognize that their role is not just to facilitate the flow of bio-data, but to act as guardians of bio-medical truth. Ensuring data integrity is not merely a technical requirement; it is a fundamental ethical imperative to prevent unstable signals from being fossilized into public health policies or insurance algorithms.

Ultimately, the thin red line of epidemiological stability marks the boundary between evidence-based medicine and data-driven theory. As we navigate the complexities of the 21st-century bioinformatic landscape, our duty is to ensure that the metrics we calculate never lose sight of the biological reality they are intended to serve.

Declarations

Acknowledgments
Not applicable.

Authors contribution
MR is the sole author of this manuscript, having conceptualized the study, performed the quantitative audit of the data, and drafted the text. The author read and approved the final manuscript.

Generative AI statement
The author declares that no Gen AI was used in the creation of this manuscript.

Conflicts of interest
The authors declares no conflicts of interest.

Ethical Approval
This research does not involve human subjects, human materials, human data, and animals. Therefore ethical approval is not necessary.

Consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
Data sharing is not applicable to this article as no new datasets were generated during the current study. All data analyzed in this manuscript are derived from the publicly available results of the cited studies.

Funding
The author received no financial support for the research, authorship, and/or publication of this article.

References
1. Weiner J, Kfuri T, Chan K, Fowles JB. (2007) "e-Iatrogenesis": The Most Critical Unintended Consequence of CPOE and other HIT. *J Am Med Inform Assoc.* May-Jun;14(3):387–388. doi: 10.1197/jamia.M2338
2. Roccetti M, Delnevo G, Casini L, et al. (2019). Is bigger always better? A controversial journey to the center of machine learning design, with uses and misuses of big data for predicting water meter failures. *J. Big Data.* 6(1):70. DOI: 10.1186/s40537-019-0235-y
3. von Elm E, Altman DG, Egger M, et al. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med*., 4(10):e296. DOI: 10.1016/j.jclinepi.2007.11.008
4 Kim H J, Kim M-H, Park SJ, Choi MG, Chun EM. (2024) Autoimmune adverse event following COVID-19 vaccination in Seoul, South Korea. *J Allergy Clin Immunol*. 2024;153(6): 1711–1720. doi: 10.1016/j.jaci.2024. 01.025
5. Kang H, Lee S. (2023). Incidence of vitiligo in South Korea: a nationwide

population-based study from 2003 to 2019. *Clin. Exp. Dermatol* 48, 484–489. doi:10.1093/ced/llad028
6. Roccetti M. (2026) Quantifying structural selection bias in observational cohort data : a ponderation analysis of age - specific incidence rates to inform vaccine safety verification. *Front Pharmacol*. 2026;16:1754809, doi: 10.3389/fphar.2025.1754809
7. Roccetti M. (2026) Logical paradoxes and epidemiological inconsistencies in vaccine-related vitiligo risk assessment. *J Allergy Clin Immunol*.. 2026 Forthcoming
8. Kim HJ, Kim MH, Choi MG, Chun EM. (2024) Psychiatric adverse events following COVID-19 vaccination: a population-based cohort study in Seoul, South Korea. *Mol Psychiatry*. 2024;29(12):3635–3643. doi: 10.1038/s41380-024-02627-0
9. Roccetti M. (2026) Before the algorithm: An exemplar case of the necessity of statistical testing for epidemiological consistency in public health data. *AIMS Public Health*. 2026;13(1):121–134. doi: 10.3934/publichealth.2026008
10. Kim HJ, Kim M-H, Choi MG, et al. (2025) 1-year risks of cancers associated with COVID-19 vaccination: a large population-based cohort study in South Korea. *Biomark Res*. 13(114). doi: 10.1186/s40364-025-00831-w
11. Park EH, Jung K-W, Park NJ, et al. (2025) Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2022. *Cancer Res Treat*. 57(2):312-330. doi: 10.4143/crt.2025.264
12. Roccetti M. (2026) A Critical Note on Contradictions in South Korean Cancer Incidence Rates: The Paradox of Crude Rates Derived from the Kim HJ et al. Cohort (*Biomark Res*, 13:114, 2025) Showing Concurrent Increases in the Vaccinated and Overall Decrease. Preprint.org Available from: https://www.preprints.org/manuscript/202510.0883
13. Cox DR. (1972). Regression Models and Life-Tables. *J R Stat Soc Series B.* 34(2):187-202. doi: 10.1111/j.2517-6161.1972.tb00899.xDigital Object Identifier (DOI)
14. Casini L, Roccetti M. (2022) Reopening Italy's schools in September 2020: A Bayesian estimation of the change in the growth rate of new SARS-CoV-2 cases. *BMJ Open*. 11(7):051458. doi: 10.1136/bmjopen-2021-051458
15. Ioannidis JP. (2005). Why most published research findings are false. *PLoS Med.* 2(8):e124. doi: 10.1371/journal.pmed.0020124
16. Greenhalgh T, Howick J, Maskrey N. (2014). Evidence based medicine: a movement in crisis? *BMJ.* 13(348):g3725. doi: 0.1136/bmj.g3725
17. Topol EJ. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 25(1):44-56. doi: 10.1038/s41591-018-0300-7
18. Kaplan RM, Irvin VV. (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Published before and after Announcement of Positive Prior Registration. *PLoS One*. 10(8):e0132382. doi: 10.1371/journal.pone.0132382
19. Pearl J. (2009). Causality: Models, Reasoning and Inference. Cambridge University Press. ISBN: 978-0-521-89560-6
20. Tosi D, Kokaj R, Roccetti M. (2024) 15 years of Big Data: a systematic literature review. *J Big Data*. 11(1):73. doi: 0.1186/s40537-024-00914-9