

Viewpoint

Is Compliance Enough? Securing STROBE with a Quantitative Stress-Test for Data Integrity

Marco Rocchetti

marco.rocchetti@unibo.it

University of Bologna, Department of Computer Science and Engineering Bologna, Italy

Corresponding Author: Marco Rocchetti

Affiliation: University of Bologna, Department of Computer Science and Engineering, Bologna, Italy

Email Address: marco.rocchetti@unibo.it

ORCID: [0000-0003-1264-8595](https://orcid.org/0000-0003-1264-8595)

Word Count: 2500

Competing Interests: The author (MR) declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Specifically:

- **Financial interests:** No financial support was received for the research, authorship, and/or publication of this article. No stocks, patents, or royalties are held in relation to the topics discussed.
- **Non-financial interests:** The author has no personal, professional, political, or institutional affiliations that could inappropriately influence the objectivity of the analysis or the proposal of the EVCC framework.
- **External relationships:** The author declares no consultancy or advisory roles for pharmaceutical companies or medical device manufacturers related to the case study presented (COVID-19 vaccines or oncology diagnostics).

Abstract

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement has established a global standard for research transparency. However, STROBE compliance does not guarantee external validity. In the era of big medical data, studies often report internally consistent associations that fail to generalize because the underlying cohorts deviate significantly from population-level baselines.

This paper proposes the External Validity Consistency Condition (EVCC), a quantitative framework designed to bridge the gap between reporting transparency and data integrity. The EVCC formalizes the relative deviation δ between cohort incidence rates and national benchmarks, providing a reproducible metric to identify selection bias.

To validate this framework, we use a recent high-profile cohort study on post-vaccination cancer risks as a test-bed. Our analysis shows that while the study is published as STROBE-compliant, its control group exhibits a 36% deficit in cancer incidence compared to national registry data. This violation of the EVCC suggests that the reported risks are likely artifacts of a non-representative sample. We argue that integrating quantitative checks like the EVCC into evidence-based medicine workflows is essential to prevent inferential overreach and ensure that public health policies are grounded in population-wide reality.

Research in context

- **Evidence before this study:** Qualitative reporting standards like STROBE focus on transparency but offer no mechanism to prevent mathematically plausible findings from unrepresentative cohorts from influencing global policy.
- **Added value of this study:** This work formalizes the transition from qualitative reporting (STROBE) to quantitative validation. We show that even massive, formally compliant datasets can harbor structural selection biases; using a recent vaccination cohort as a proof-of-concept, we reveal a representativeness gap that fundamentally undermines the reliability of its public health conclusions.
- **Implications of all the available evidence:** The EVCC provides a scalable, reproducible stress-test that should be integrated into evidence appraisal workflows. It enables journal editors and health agencies to differentiate between population-wide signals and cohort-specific artifacts, thereby safeguarding public health discourse from inferential overreach.

Keywords: *STROBE Statement, External Validity Consistency Condition (EVCC), Inferential Overreach, Data Integrity, Quantitative Evidence Appraisal, Research Quality Gatekeeping.*

Introduction

Observational studies are indispensable when randomized trials are infeasible, yet they remain intrinsically vulnerable to confounding, selection effects, and limits on generalizability. Reporting guidelines such as STROBE were introduced to mitigate these risks by making the boundaries of legitimate inference explicit [1]. Increasingly, however, formal STROBE compliance appears to function as a procedural requirement rather than as an epistemic constraint.

The STROBE statement in fact has fundamentally transformed the landscape of observational research by establishing rigorous standards for reporting. By mandating detailed descriptions of study designs, participants, and statistical methods, it has mitigated many forms of reporting bias. However, a critical gap remains.

STROBE is a qualitative checklist, not a quantitative validator. In the current era of data sprawl, where massive electronic health records and administrative datasets are repurposed for epidemiological research, the internal consistency of a model (internal validity) is often mistaken for its applicability to the general population (external validity).

This often leads to inferential overreach, where statistically significant associations found in a specific, often skewed, cohort are erroneously presented as universal truths. There is an urgent need for a formal, data-driven guardrail that allows researchers, editors, and policy-makers to verify a cohort's representativeness before its findings influence clinical practice.

We propose the External Validity Consistency Condition (EVCC) as a useful quantitative complement to the qualitative framework of STROBE.

Methods

In this Section, we provide all the necessary details on methods used in this study, allowing readers to easily replicate our findings.

The EVCC Mathematical Framework

The core objective of the External Validity Consistency Condition (EVCC) proposal is to quantify the distance between a study cohort and the population it claims to represent.

We formalize this through the concept of relative deviation δ . Let $CR(Study)$ be the crude incidence rate of a specific outcome within the study cohort, and $CR(Pop)$ be the corresponding rate in the target population derived from authoritative national benchmarks. The deviation between study cohort and population is defined as:

$$\delta = |CR(Study) - CR(Pop)| / CR(Pop) . \quad (1)$$

We propose that a cohort satisfies the External Validity Consistency Condition (EVCC) only if:

$$\delta = O(\varepsilon) , \quad (2)$$

where ε represents a small, acceptable threshold of tolerance (typically 0.05 or 5%). Using the $O(.)$ notation emphasizes that EVCC constrains the order of magnitude of divergence, not just a single numeric instance. When δ significantly exceeds this order of magnitude, the cohort exhibits a systematic divergence that invalidates population-level inferences.

This formalism generalizes EVCC to any observational dataset, providing conceptual rigor for meta-research applications. Values substantially exceeding ε indicate systematic divergence, signaling that population-level inferences would be unreliable. This approach allows for a transparent assessment of external validity using only published summary statistics.

The Graduated Scale of δ : From Stochastic Noise to Structural Failure

The strength of the EVCC lies in its ability to quantify the reliability of a cohort through a graduated scale of divergence. It moves beyond a simple valid/invalid binary to offer a nuanced appraisal of medical data integrity, defining the following scenarios:

- Scenario A: $\delta \leq \varepsilon$ (The Safe Zone). The cohort incidence mirrors the population benchmark within a minimal margin of error (e.g., 5%). The cohort is reasonably representative. Findings are highly generalizable and the risk of selection bias affecting the final estimates is negligible.
- Scenario B: $\varepsilon < \delta < 0.15$ (The Cautionary Zone). There is a moderate divergence (e.g., 10% to 15%). The cohort is not a perfect proxy, but it is not a statistical outlier. Hence, researchers just need to provide a transparent justification for the divergence (e.g., specific age-group recruitment). Sensitivity analyses can be useful to ensure that this moderate drift does not flip the direction of the reported associations.
- Scenario C: $\delta \gg \varepsilon$ (The Danger Zone). The deviation is an order of magnitude higher than the tolerance threshold. We are here in presence of a critical inferential overreach. Deficits of the order of 20/30/40% or more indicate a structural failure in representativeness. The cohort is systematically different from the target population, rendering relative risk estimates (like Hazard Ratios) likely artifacts of sampling rather than biological signals.

Visualizing the Integrity Gradient

To implement the EVCC as a strategic policy tool, we propose a kind of methodological traffic light system that can be integrated into automated peer-review and Evidence-Based Medicine (EBM) workflows as represented in Table 1 below.

Table 1: Methodological Traffic Light System

δ Value	Signal	Strategic Recommendation
≤ 0.05	Green (Safe)	Proceed with inference; high generalizability to the target population
$0.05 < \delta < 0.15$	Yellow (Caution)	Flag as Cohort-Specific findings: requiring deeper audit of exclusion criteria
≥ 0.15	Red (Danger)	Evidence of Critical Bias. Results should not inform broad public health policy

By adopting this gradient, the EVCC does not demand a utopian $\delta = 0$. Instead, it serves as a calibrated audit. While minor fluctuations are expected in real-world data, a deviation of some 0.40, for example, would represent a catastrophic breakdown of external validity that no internal statistical fine-tuning or Propensity Score Matching procedure can correct [3, 4].

A Real Case Study as a Test Bed

To validate the EVCC, we analyzed the summary statistics from a large-scale South Korean study [2], which investigated the 1-year risk of various cancers following COVID-19 vaccination. This study represents a perfect candidate: it is massive (approximately 3 million), utilizes sophisticated propensity score matching, and is formally STROBE-compliant.

We compared the study’s internal cancer incidence rates, reported in Table 2, against the South Korean Central Cancer Registry (KCCR) data, as reported in the official literature [5-7], which provides the gold-standard annual crude incidence for the South Korean population during the 2020–2022 period (Table 3).

Cohort Group	Observed Cases	Cohort Size	CR(Study)
Overall cohort	12,133	2,975,035	40.78
Vaccinated Group	10,144	2,380,028	42.63
Unvaccinated (Control)	1,989	595,007	33.43

Table 2: Summary Statistics from the South Korean Cohort [2].

Year	CR(Pop) per 100,000	CR(Pop) per 10,000
2020	482.9	48.29
2021	540.6	54.06
2022	550.2	55.02

Table 3: Official National Crude Incidence Rates (CR) for All Cancers in South Korea per 100,000 and the derived $CR(Pop)$ per 10,000, used to establish the national average baseline for the reference period (2020–2022).

Consequently, the official average $CR(Pop)$ baseline for all cancers for the reference period (2020–2022) can be established (from Table 3) as the mean of these values: $CR(Pop)$ (Official Average) = 52.46 per 10,000 (Standard Deviation, $SD = 2.97$).

Patient and Public Involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Results

The analysis of the raw data from the cohort of Table 2 immediately highlights a significant internal-external discrepancy when processed through the EVCC framework (Table 4).

Analysis Group	CR(Study)	CR(Pop)	Deviation δ	EVCC Status
Overall cohort	40.78	52.46	0.22	Violated
Vaccinated	42.63	52.46	0.19	Violated
Unvaccinated	33.43	52.46	0.36	Critical Violation

Table 4: EVCC Verification against National Benchmarks (KCCR).

The results indicate that the unvaccinated control group exhibits a 36% deficit in cancer incidence ($\delta = 0.36$) compared to the national average. Such a magnitude of divergence is mathematically incompatible with the claim of a representative sample.

Discussion

The subsequent Discussion analyzes our findings, focusing on several areas: 1) summarizing the results from the investigated case study [2]; 2) discussing the limitations of our proposal; 3) implementing EVCC in EBM medicine scenarios.

Selection Bias and the Healthy User Artifact

The violation of the EVCC in the South Korean test-bed uncovers a profound selection bias. An incidence rate in the control group that is one-third lower than the national average suggests that the non-vaccinated pool was systematically healthier or younger than the general population, a classic Healthy User Bias [8].

When the baseline risk in the control group is artificially suppressed, any comparison with a more normal vaccinated group will yield an inflated Hazard Ratio (HR). This creates a statistical mirage: the study detects an increased risk not because of a biological effect of the vaccine, but because the denominator (the control group) is missing expected cancer cases [9].

Limitations of the EVCC Framework

Despite its utility, the EVCC is not a panacea and has specific limitations.

First, its accuracy depends entirely on the quality of the population benchmark. If national registry data are outdated or under-reported, the δ calculation will be flawed.

Second, the EVCC is a crude diagnostic tool: it flags a representativeness problem but does not, by itself, identify the source of the bias (e.g., whether it is due to age distribution, socioeconomic factors, or matching artifacts).

Third, the threshold ε is conceptually flexible; while we suggest 0.05, different clinical fields might require more or less stringent thresholds depending on the outcome's volatility.

In closing, the EVCC should therefore be viewed as an early warning system, a first-line quantitative test that, if failed, mandates a deeper investigation into the cohort's composition.

Implementing EVCC in Evidence-Based Medicine

We argue that the EVCC should be integrated into the appraisal workflow of Evidence-Based Medicine (EBM). If a study violates the EVCC condition: $\delta \gg \varepsilon$, its results should be clearly labeled as cohort-specific rather than population-generalizable. This distinction is vital for policy-makers who must decide whether to act on observational findings.

Conclusion

The proposed EVCC framework serves as a critical epistemological bridge between the qualitative transparency mandated by STROBE [1] and the objective necessity of data integrity in the Big Data era. While reporting standards have successfully mitigated publication bias, they remain insufficient against the more insidious threat of structural selection bias, which often remains hidden behind the veil of massive sample sizes and sophisticated algorithmic adjustments.

Our application of the EVCC to a high-profile vaccination cohort [2] serves not merely as a critique of a specific study, but as a proof-of-concept for a systemic vulnerability. It demonstrates that even the most formally compliant and computationally advanced datasets can fail the test of external reality when they lose their anchor to population-level benchmarks. This statistical mirage highlights a broader risk: without quantitative guardrails, the very tools intended to advance public health can inadvertently fuel inferential overreach, translating cohort-specific artifacts into global policy.

In an age where medical decisions are increasingly data-driven, the scientific community must transition from a culture of "transparency as compliance" to one of "integrity as auditability". By adopting the EVCC as a standard stress-test, journal editors, regulatory bodies, and researchers can safeguard public health discourse, ensuring that the evidence-based medicine of the future is grounded not just in the volume of data, but in its representativeness and truth.

Contributors: MR as a single author conceived, designed, wrote, managed, and revised this manuscript, and has read and agreed to the published version of the manuscript.

Funding: The author does not declare a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Research Ethics Approval: Not applicable, neither humans nor animals nor plants nor personal data were involved in this study.

Data availability statement: Data sharing not applicable as no datasets generated and/or analyzed for this study. Further reasonable requests can be addressed to the corresponding author (email: marco.rocchetti@unibo.it).

Patient consent for publication; Not required.

Transparency: The sole author (MR) affirms that the manuscript is an honest, accurate, and transparent account of the computational and methodological analysis being reported; that no important aspects of the publicly available data and calculations have been omitted; and that all findings are derived exclusively from the cited public data sources, ensuring the total reproducibility of the study.

Artificial Intelligence Statement: The author affirms that all original ideas, final data interpretations, linguistic refinement of the ideas, and the definitive conclusions are exclusively product of his own intellectual responsibility without use of AI tools.

References

- 1 E. Elm, D.G. Altman, M. Egger, et al., The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med.* 4(10):e296. (2007) DOI: 10.1016/j.jclinepi.2007.11.008
- 2 Kim HJ, Kim M-H, Choi MG, et al. 1-year risks of cancers associated with COVID-19 vaccination: a large population-based cohort study in South Korea. *Biomark Res.*, 13(114). (2025) doi: 10.1186/s40364-025-00831-w
- 3 Murad MH, Katabi A, Benkhadra R, et al. External validity, generalisability, applicability and directness: a brief primer. *BMJ Evid Based Med*, 23:17-19. (2018) doi: 10.1136/ebmed-2017-110800.
- 4 Wijn SRW, Rovers MM, Hannink G. Confounding adjustment methods in longitudinal observational data with a time-varying treatment: a mapping review. *BMJ Open*, 12:e058977. (2022) doi: 10.1136/bmjopen-2021-058977
- 5 Kang MJ, Jung K-W, Bang SH, et al. Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2020. *Cancer Res Treat.*, 55(2):385-399. (2023) doi: 10.4143/crt.2023.447
- 6 Park EH, Jung K-W, Park NJ, et al. (Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2021. *Cancer Res Treat.*, 56(2):357-371. (2024) doi: 10.4143/crt.2024.253
- 7 Park EH, Jung K-W, Park NJ, et al. Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2022. *Cancer Res Treat.*, 57(2):312-330. (2025) doi: 10.4143/crt.2025.264
- 8 Hernan MA, Hernandez-Diaz S, Robins JM, A structural approach to selection bias. *Epidem.*, 15(5):615-625. (2004) doi: 10.1097/01.ede.0000135174.63482.43
- 9 Roccetti M, 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. Preprints.org. (2025) doi: 10.20944/preprints202510.0883.v1