
Addendum to “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)” - On Possible Sampling Bias and Inverted Propensity Score Matching

[Marco Roccetti](#) *

Posted Date: 22 October 2025

doi: 10.20944/preprints202510.1664.v1

Keywords: Propensity Score Matching; Inversion or Misapplication of the 1:4 Matching; Crude Incidence Rate; Epidemiological Paradox; COVID-19 Vaccination; Cancer Incidence



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Short Note

Addendum to “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)” - On Possible Sampling Bias and Inverted Propensity Score Matching

Marco Rocchetti

Department of Computer Science and Engineering, University of Bologna, Italy; marco.rocchetti@unibo.it

Abstract

This addendum proposes a working hypothesis to explain the epidemiological paradox identified in our original note [1] regarding the crude cancer incidence rates reported by Kim HJ et al. (2025) [2]. Specifically, the disproportionate size of the COVID-19 vaccinated and unvaccinated groups in the study's final matched cohort suggests that the stated 1:4 propensity score matching (PSM) ratio may have been inverted or misapplied. Instead of selecting four unvaccinated controls per vaccinated individual, the study cohort contains approximately four vaccinated individuals per unvaccinated control. This inversion of treatment and control groups could bias the study's incidence estimates and contribute to the paradoxical observation that the vaccinated group shows higher cancer incidence while the overall crude incidence rate in the cohort is lower than national averages. Clarifying the PSM methodology is essential to resolve this contradiction and validate the study's conclusions.

Keywords: Propensity Score Matching; Inversion or Misapplication of the 1:4 Matching; Crude Incidence Rate; Epidemiological Paradox; COVID-19 Vaccination; Cancer Incidence

1. Introduction

In the original critical note [1], a paradox emerged from the analysis of crude cancer incidence rates (CRs) reported by Kim et al. [2]. While the vaccinated subgroup in the cohort exhibited a higher incidence of new cancer cases compared to the unvaccinated, the overall cohort's crude incidence rate was substantially below the official national cancer incidence reported by Korean registries.

More precisely: despite the study [2] reporting a higher crude cancer incidence rate (CR) in vaccinated individuals (42.63 per 10,000) than unvaccinated (33.43 per 10,000), the overall cohort's CR (40.78 per 10,000) was markedly lower than South Korea's official national average CR (~52.46 per 10,000). This discrepancy suggested a representativeness problem.

This addendum builds on that observation by examining the reported propensity score matching procedure in the study, hypothesizing that an inversion or misapplication of the 1:4 matching ratio may underlie the paradox.

2. Data Summary and Observed Matching Discrepancy

Kim et al. report a final matched cohort of 2,975,035 individuals comprising 2,380,028 vaccinated and 595,007 unvaccinated individuals [2]. This distribution corresponds to approximately four vaccinated individuals for every unvaccinated individual:

2,380,028 (vaccinated) : 595,007 (unvaccinated) $\approx$ 4:1

By definition, Propensity score matching aims to reduce confounding by matching treated individuals with comparable untreated controls. In vaccine effectiveness or adverse effect studies, the natural and appropriate approach is to match each COVID-19 vaccinated individual (treated) with one or more unvaccinated individuals (controls). This ensures that the control group is constructed relative to the treated group.

However, as already explained, given the cohort sizes reported, vaccinated individuals (2,380,028) outnumber unvaccinated (595,007) by a ratio close to four. A 1:4 PSM would typically imply matching each unvaccinated individual to four vaccinated controls, yet the reported data show the opposite: 595,007 unvaccinated and 2,380,028 vaccinated.

This suggests that the matching may have been performed “in reverse” i.e., the control group (unvaccinated) was constructed by matching vaccinated individuals rather than the other way around. Such inversion would produce a matched cohort not representative of the true underlying population structure.

3. Implications of Matching Inversion

The likely inversion or misapplication of the matching procedure raises multiple methodological concerns:

- **Incorrect Group Assignment:** Vaccinated individuals should represent the treatment group, with unvaccinated individuals serving as controls. Reversing this order undermines the matching's purpose.
- **Underrepresentation of Unvaccinated Population:** If controls (unvaccinated) are chosen based on vaccinated individuals, the smaller unvaccinated pool is overextended, potentially selecting unrepresentative samples.
- **Biased Effect Estimates:** Hazard ratios and incidence rate comparisons based on improperly matched groups may be invalid, leading to unreliable conclusions about vaccination's association with cancer incidence.
- **Cohort Representativeness:** The inflated vaccinated group and smaller unvaccinated group suggest sampling bias, reducing external validity and potentially skewing incidence estimates downward, consistent with the observed paradox.
- **Bias in Cancer Incidence Estimates:** The inverted matching direction could lead to artificially reduced crude incidence rates overall, explaining the paradoxical discrepancy with national averages.

4. Relation to the Epidemiological Paradox

This hypothesis offers a plausible explanation for the paradox in crude incidence rates noted in the original analysis. The paradoxical pattern, higher cancer incidence in vaccinated individuals yet an overall cohort incidence rate below national averages, may stem from flawed group construction in PSM, affecting the representativeness and internal consistency of the cohort.

Confirming this hypothesis requires access to the underlying dataset and detailed methodology, currently unavailable. We hence reiterate the call for public access to the Korean National Health Insurance database used by Kim et al. [1], to enable independent validation and clarify these methodological issues.

5. Conclusion and Recommendations

To clarify these concerns and reinforce scientific rigor, the following actions are essential:

- **Transparency from Authors:** A detailed description of the matching methodology, including algorithmic details and group assignment, should be disclosed by Kim et al.
- **Data Access:** Public availability of the underlying Korean National Health Insurance database would allow independent verification and alternative analyses.

- **Rigorous Review:** Proper application and reporting of PSM in observational vaccine safety studies are critical to avoid misleading interpretations.

Resolving the identified methodological ambiguity is fundamental to trust in the reported associations and broader vaccine safety assessments.

Funding: This research received no external funding.

Data Availability Statement: The data presented here is either included directly or was extracted from the referenced documents. All calculations are easily reproducible based on the definitions provided.

Ethics approval and consent to participate: This study uses publicly available, aggregated data that contains no private information. Therefore, ethical approval is not required.

Consent for publication: Not applicable.

Conflicts of Interest: The author declares no competing interests.

References

1. Rocchetti 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
2. Kim HJ, Kim M-H, Choi MG, Chun EM. 1-year risks of cancers associated with COVID-19 vaccination: a large population-based cohort study in South Korea. Biomark Res, 13, 114 (2025). <https://doi.org/10.1186/s40364-025-00831-w>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.