

Original Research

Quantifying Structural Selection Bias in Observational Cohort Data: A Ponderation Analysis of Age - Specific Incidence Rates to Inform Vaccine Safety Verification

Marco Rocchetti

Department of Computer Science and Engineering

University of Bologna, 40126, Italy

marco.rocchetti@unibo.it

ORCID: 0000-0003-1264-8595

Abstract

Background: A recent nationwide cohort study reported an unadjusted Hazard Ratio (HR) of 2.714 for Vitiligo incidence following COVID-19 vaccination, indicating a major safety concern. This finding was based on cohorts with an ≈ 11 -year age difference, immediately raising critical concerns regarding extreme structural selection and detection bias.

Objectives: We hypothesize that this extreme association is an artifact of a fatal methodological flaw, challenging the study's internal validity and subsequent external validity. We aim to quantitatively separate the HR attributable to the structural age imbalance (HR Structural) from the residual HR (HR Residual), which quantifies the uncorrected methodological failure. We further perform a plausible recalculation of risk to demonstrate the complete collapse of the risk signal upon correcting the methodological failure in the baseline cohort.

Methods: We performed a stratified ponderation analysis using the age distribution of the scrutinized study's cohorts (Vaccinated, mean age=56.32 years vs Non-Vaccinated, mean age=45.51 years) and applied established national age-specific Vitiligo incidence rates (IR) from external epidemiology.

Results: The HR Structural was calculated to be 1.21. The remaining HR Residual of 2.24 quantifies the uncorrected methodological failure. The NV cohort's observed incidence rate (0.67/10,000) was found to be nearly 70% lower than the expected rate (2.21 / 10,000), providing quantifiable evidence of profound non-comparability. The subsequent recalculation of risk, correcting for this baseline failure, reduces the observed HR of = 2.714 to an HR Corrected of almost a unity, thus completely annulling the signal of risk due to vaccination.

Discussion: The HR= 2.714 of the scrutinized study is an unstable statistical artifact. The overwhelming majority of the observed association is a consequence of a fatal design flaw. The HR Corrected of almost 1 confirms that correcting the methodological error eliminates the risk signal, demonstrating a severe lack of internal and external validity of the original study.

Keywords: Covid-19 vaccine safety, Structural selection bias, Ponderation analysis, Hazard ratio decomposition, Internal/External validity, Detection bias

1. Introduction

Observational studies utilizing national registries, such as those conducted in South Korea [1], represent a critical resource for post-marketing surveillance and vaccine safety verification. However, the reliance on pre-existing data necessitates strict adherence to established methodological standards, notably the STROBE (STrengthening the Reporting of Observational studies in Epidemiology) guidelines [2]. The primary goal is to ensure internal validity, that the observed association is real within the study context, which is a prerequisite for achieving external validity (that is generalizability to the broader population).

The recent study published in [1] reported a strikingly high, unadjusted Hazard Ratio (HR_{Gross}) of 2.714 for Vitiligo following COVID-19 vaccination, based on a comparison between a Vaccinated (V) cohort (mean age=56.32 years) and a Non-Vaccinated (Non-Vaccinated) cohort (mean age=45.51 years). This ≈11-year age difference immediately flagged critical concerns regarding confounding by indication and immortal time bias [3]. The sheer magnitude of the ≈11-year age difference, coupled with the cumulative incidence rates observed (2.22 vs 0.67 per 10,000), strongly suggests that the cohorts were inherently non-comparable.

Our analysis posits that the reported $HR=2.714$ is not a reflection of a robust biological signal but rather a quantitative measure of a fatal design flaw. We hypothesize that an Extreme Structural Selection and Detection Bias was introduced by defining the cohorts in a manner that artificially minimized the baseline risk in the NV group, while simultaneously maximizing the detection and prevalence risk in the V group. We present a rigorous, quantitative method, that is a stratified ponderation analysis using external South Korean national age-specific incidence data, to decompose the observed HR and isolate the true contribution of the structural bias [6-10].

The quantitative findings of the present research confirm this hypothesis: the structural age difference alone accounts for a calculated Structural Hazard Ratio (HR Structural) of 1.211. Importantly, we demonstrate that the non-comparability of the baseline cohort (NV) caused its observed incidence rate to drop nearly 70% below the expected rate. The overwhelming majority of the observed association is captured by the residual Hazard Ratio (HR Residual of 2.24) which stands as a clear measure of the uncorrected methodological failure. We further show that applying a reasonable recalculation of risk, which corrects the failure of the baseline incidence rate, reduces the observed risk signal from 2.714 to a **clinically insignificant** HR of approx. 1 (HR Corrected).

Essentially, the quantitative findings of the present research has move along two concatenated directions. First we confirm this hypothesis that the structural age difference alone accounts for a calculated structural Hazard Ratio (HR Structural) of 1.21. This means that the observed demographic imbalance explains more than the 12% of the initially reported excess risk. Then, we show that the majority of the association captured by the residual Hazard Ratio (HR Residual) of 2.24 simply stands as a clear measure of the uncorrected methodological failure. This substantial residual value not only strongly indicates that the cohorts were not subject to a common support, leading to a profound violation of the assumption of comparability required by the Cox Proportional Hazards model but can be also explained if one takes into account the drop of nearly 70% in the incidence rate of Vitiligo in the NV group. Putting this information into the calculation easily leads to a corrected value of HR of almost 1 that practically annuls the risk difference between V and NV subgroups.

Ultimately, the goal of this re-evaluation is to reassert the imperative for epidemiological validity in studies of vaccine safety derived from observational data. We demonstrate that sophisticated statistical adjustments cannot remedy fundamental flaws in cohort design where non-measured confounding factors, such as health-seeking

behavior, surveillance frequency and distribution of the disease incidence peaks are unevenly distributed [7]. By quantitatively isolating and measuring the non-causal structural bias, our analysis provides a critical framework for interpreting extreme risk estimates and ensuring that public health conclusions are based on associations that are epidemiologically sound, rather than artifactual.

2. Methods

We here provide all the fundamental methods and data useful for the aim of pondering the structural bias on which we are investigating.

2.1 Study Data and Baseline Characteristics

We extracted the following key data from [1] to establish the basis of the structural bias as reported in the following Table 1.

Table 1: Baseline Characteristics and Unadjusted Incidence Rates from [1].

Cohort	Mean Age	Standard Deviation (SD)	Cumulative Incidence Rate (at 3 mo) (per 10,000 p-y)
Non-Vaccinated (NV)	45.51 years	17.31	$P(NV) = 0.67$
Vaccinated (V)	56.32 years	16.55	$P(V) = 2.22$

2.2 Stratified Ponderation Analysis

We performed a first preliminary quantitative analysis by combining the age distribution percentages $P(i)$ of the V and NV groups of [1] with independent, established age-specific annual incidence rates $IR(i)$ for Vitiligo in South Korea, based on 2019 data as reported in [5]. It is reported in Table 2 below.

Table 2: Input Data for Ponderation Analysis

Age Group	South Korean IR (per 10,000 p-y)	% in NV Group P(i, NV)	% in V Group P(i,V)
< 20 y	3.4241	Not included in [1]	Not included in [1]
20-29 y	1.5717	18.46%	9.92%
30-39 y	1.7813	25.49%	7.70%
40-49 y	1.9053	20.92%	10.82%
50-59 y	2.5874	14.16%	24.76%
>= 60 y	3.3643	20.97%	45.79%
Total	—	100%	100%

A key observation is that, it should be noted here that the composition of the NV group is strongly influenced by the presence of young individuals who are already past the first peak of Vitiligo onset, contrasting with the massive presence in the NV group of older individuals entirely included in the interval where the second incidence peak of Vitiligo manifests, which is universally recognized as having a bimodal distribution (under 20, over 40/50 years) [5].

2.3 Calculation of HR Structural and HR Residual

The Expected Annual Incidence Rate, IR(Expected,) for each cohort, based solely on its structural age composition, can be calculated using the following Formula 1:

$$IR(\text{Expected}) = \sum (IR(i) \times P(i)). \quad (1)$$

Where $IR(i)$ are the South Korean age-specific incidence rates from external data (Table 2) and $P(i)$ are the proportional distributions of the respective cohort (V or NV) reported in the same Table 2.

Applying this ponderation to the Non-Vaccinated (NV) cohort demographics, we obtain the baseline expected incidence, $IR(NV, \text{Expected})$ exactly as follows:

$$IR(NV, \text{Expected}) = (1.5717 \times 0.1846) + (1.7813 \times 0.2549) + (1.9053 \times 0.2092) + (2.5874 \times 0.1416) + (3.3643 \times 0.2097) \approx 2.2146 / 10,000.$$

Similarly, applying the ponderation to the Vaccinated (V) cohort demographics yields $IR(V, \text{Expected})$:

$$IR(V, \text{Expected}) = (1.5717 \times 0.0992) + (1.7813 \times 0.0770) + (1.9053 \times 0.1082) + (2.5874 \times 0.2476) + (3.3643 \times 0.4579) \approx 2.6804 / 10,000.$$

This allowed us to calculate the HR Structural as follows: $HR \text{ Structural} = 2.6804 / 2.2146 \approx 1.21$.

Finally, the HR Residual can be computed as the ratio between the HR provided in [1] (termed HR Observed) and our computed HR Structural: $HR \text{ Residual} = HR \text{ Observed} / HR \text{ Structural} = 2.714 / 1.21 \approx 2.24$.

The consistency of this multiplicative decomposition is confirmed by the near-exact reconstruction of the observed value: $HR \text{ Structural} \times HR \text{ Residual} \approx 1.21 \times 2.24 \approx 2.714$.

3. Results

We here provide the two main set of results stemming from our analysis whose methodology has been anticipated in the previous Section.

3.1 Decomposition of the Observed Hazard Ratio and Incidence Discrepancy

Following the calculation of the Expected Incidence Rates $IR(\text{Expected})$ based solely on the structural age compositions of the two cohorts (Section 2.3), we proceeded to quantify the true extent of the methodological failure. This involved decomposing the high, observed $HR \text{ Observed} = 2.714$ from [1] into two distinct components: the risk attributable purely to the structural age imbalance ($HR \text{ Structural}$) and the risk stemming from all other

uncorrected design flaws and selection biases (HR Residual). Since Hazard Ratios combine *multiplicatively*, that is $HR_{Observed} = HR_{Structural} \times HR_{Residual}$, the HR Residual thus acts as a precise metric for the degree of non-comparability that persists despite accounting for the known age difference. The breakdown of this risk is definitely presented in Table 3.

Table 3: Decomposing the Observed Hazard Ratio (HR=2.714 [1])

Parameter	Description	Value	Contribution to Excess Risk: $(HR_x - 1) / (HR_{Obs} - 1) \times 100$
HR Observed	Unadjusted Hazard Ratio from [1]	2.714	100%
HR Structural	HR due to Age Structural Bias Alone	1.21	12.25%
HR Residual	HR Unexplained by Structural Age Bias	2.24	72.35%*

The most salient finding of the ponderation analysis shown above is the extreme divergence between the expected and observed incidence rates in the reference cohort. The demographic composition of the Non-Vaccinated (NV) subgroup predicted an expected Incidence Rate of 2.21 / 10,000 based on established national age-specific rates. However, the rate observed in [1] was only 0.67 / 10,000. This discrepancy means the observed incidence in the reference cohort was nearly 70% lower than methodologically expected, providing immediate, quantifiable evidence of the profound non-comparability of the baseline groups.

Furthermore, it is crucial to emphasize how our robust, age-specific ponderation analysis has shown that the structural age difference explains only 12.25% of the excess risk signaled by the authors of [1]. The overwhelming majority of the association (72.35%, resulting in an HR Residual of ≈ 2.24) is entirely attributable to uncorrected methodological flaws which should be attributed to a basic failure in the construction of the cohort and their

subgroups as hypothesized below. It is also to be noted (leftmost bottom cell of the Table above) that being this modeling multiplicative in nature its translation in percentages hides a so called Interaction Term, mostly due to rounding during the relative operations. This explain why we do not get a perfect 100%.

3.2 Recalculation of Risk: The Collapse of the Observed Association

The massive quantitative discrepancy identified in Section 3.1, where the $IR(NV, \text{Observed})$ of $0.67 / 10,000$ falls short of the $IR(NV, \text{Expected})$ of approx. $2.21 / 10,000$, by nearly 70%, due to non-comparability bias, allows us to perform a fundamental recalculation of risk. We obviously hypothesize that had the NV cohort been properly matched for unmeasured confounders, the true baseline risk would have approached the expected incidence rate of $2.21 / 10,000$. Furthermore, if we assume, for the sake of this calculation, that the Detection Bias acting on the Vaccinated (V) cohort is negligible (i.e., that $IR(V, \text{Observed}) \approx IR(V, \text{True})$), we can calculate the resulting Hazard Ratio (HR Corrected) as follows: $HR \text{ Corrected} = IR(V, \text{Observed}) / IR(NV, \text{Expected}) \approx 1$.

This simple recalculation demonstrates that by correcting the methodological failure of the denominator alone, the observed risk signal of 2.714 collapses entirely to ≈ 1 . An HR of 1 would mean a clinically insignificant increased risk, effectively reducing the study's finding to a null hypothesis.

4. Discussion

We will summarize here the key takeaways of this discussion into two primary issues to ensure they are properly highlighted, noting that at the heart of the matter lie problems of loss of comparability and resulting clinical significance.

4.1 The Collapse of Internal Validity: The Double Distortion Mechanism

The persistence of the high residual HR (2.24) after robust adjustment for age structure ($HR \text{ Structural} = 1.21$) provides a definitive quantitative proof that the cohorts have been constructed as non-comparable. The study's design of [1] suffers from a double distortion mechanism that fundamentally violates the core premise of observational epidemiology.

First, the artificial baseline depression of the NV sub-group is attributed to the Selection Bias revealed in Section 3.1: the IR(NV) being 70% lower than expected indicates that the individuals composing the denominator are not representative of the general population. This massive deficit is caused by the inclusion of a minority of older individuals (≥ 60 years) who are part of an exceptionally healthy survivor cohort and/or have minimal medical interaction [8], thus ensuring profound under-detection of Vitiligo cases. This failure to capture the true baseline risk, depressing the denominator from the expected 2.21 / 10,000 to the observed 0.67 / 10,000, is the primary quantitative driver of the resulting high HR Residual ≈ 2.24 .

Second, we confront an inflated incidence by detection and risk in the V subgroup. Conversely, the V group's composition of [1] ($\approx 70\%$ aged ≥ 50 years) guarantees maximal risk exposure because this demographic actively intercepts the second, higher incidence peak of Vitiligo (as seen in the Incidence Rate data before). This structural flaw, coupled with the heightened surveillance and utilization bias typical of vaccinated groups, ensures that both the intrinsic risk and the diagnostic detection rate are maximized. This extreme structural separation, especially in the high-risk and high-surveillance age categories, represents a violation of the common support assumption. The Cox model employed in the scrutinized study, therefore, did not compare like with like, but rather measured the risk differential between an artificially clean control group and a maximally surveilled risk group.

4.2 Clinical and External Validity Implications for Vaccine Safety Verification

The extreme HR Residual ≈ 2.24 cannot be interpreted as a genuine biological effect. The recalculation of Risk performed in Section 3.2, which demonstrates the complete collapse of the risk signal to $HR \approx 1$ upon correcting the methodological baseline failure, provides compelling evidence against a genuine biological association. If the HR Residual were indeed the genuine vaccine effect, it would imply that the extensive methodological failure observed in the denominator ($\approx 70\%$ under-reporting) must be entirely compensated by an unmeasured bias in the numerator. Epidemiologically, it is far more plausible that the 2.24 represents the cumulative effect of uncorrected Selection and Detection Bias, a direct consequence of the study's faulty design.

The clinical implication is that the reported $HR = 2.714$ of [1] is gravely misleading for patients and clinicians. It does not reflect the incremental risk of vaccination but rather the difference in underlying health and healthcare seeking behavior between two demographically distinct groups in South Korea. This methodological failure is a

serious breach of epidemiological reporting standards in the sense of the STROBE protocol and undermines the utility of national registry data for assessing vaccine safety signals when proper cohort matching is neglected.

4.3 Limitations and Future Directions

We acknowledge several limitations to our ponderation analysis. First, the HR Structural calculation relies on the assumption that the external, age-specific incidence rates $IR(i)$ derived from the general South Korean population (as reported from different perspectives in all the available literature [4-6]) accurately reflect the true baseline risk within the national health insurance service data utilized in [1]. Second, our analysis only addresses confounding introduced by structural age differences; we are unable to quantify the residual contributions of other unmeasured variables, such as socioeconomic status (SES), co-morbidities, or the precise effect of Detection Bias related to varying healthcare utilization frequency, all of which likely inflated the HR Residual. Furthermore, this methodological criticism does not exclude the possibility of a smaller, genuine biological signal, which would be revealed only through a properly designed study utilizing tight Propensity Score Matching (PSM) and time-varying exposure analysis which was clearly not used by the authors of [1]. Nonetheless, we maintain that our present analysis has provided a crucial quantitative framework for critically assessing the validity of large epidemiological risk estimates derived from imbalanced cohorts and providing a relevant contribution towards the fidelity and verifiability of vaccine safety signals derived from observational cohort data.

5. Conclusion

The association between COVID-19 vaccination and Vitiligo ($HR = 2.714$) as reported in [1] is an extreme statistical artifact. Our robust ponderation analysis, based on specific South Korean age-incidence rates, definitively proves that the structural age imbalance explains only a minor fraction ($HR_{\text{Structural}} \approx 1.21$) of the observed risk. The overwhelming HR Residual of 2.24 is the quantitative measure of the methodological failure caused by the Structural Selection and Detection Bias that have affected the construction of the retrospective cohort of [1]. This failure to establish genuinely comparable cohorts has compromised both the internal validity and external validity of the investigated study. The reported finding of [1] is therefore an association primarily driven by bias (a spurious association) and should not be used to inform public health policy or safety communication. Our attempt to recalculate the real risk demonstrates that correcting the denominator's failure results in an $HR \approx 1$, which is a plausible outcome consistent with the fundamental rules of biostatistics required

for a valid cohort study. We conclude by urging the re-evaluation and potential reconsideration of the study's conclusions. Our findings also underscore the persistent reliance on and respect for methodological goldstandards in clinical research [9, 10]. While innovation in epidemiological design is crucial, these established benchmarks must only be challenged or superseded by new studies featuring superior internal validity and robust correction for all known sources of bias.

Author Information

Marco Rocchetti: Department of Computer Science and Engineering, University of Bologna, 4016 Bologna, Italy, marco.rocchetti@unibo.it. ORCID: 0000-0003-1264-8595, sole and corresponding author

Author Contributions

MR conceived and designed the study, carried out all data collection and analysis, interpreted the quantitative results, and was the sole author responsible for writing and revising the manuscript. The author affirms full responsibility for the integrity of the data and the accuracy of the data analysis presented.

Ethics approval and consent to participate

This study uses publicly available, aggregated data that contains no private information. Therefore, ethical approval is not required

Data Availability Statement

The data presented here is either included directly or was extracted from the referenced documents and cited literature. All calculations are easily reproducible based on the definitions provided.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. This study was conducted entirely independently by the author using personal resources.

Conflict of Interest

The author declares that there is no conflict of interest, financial, personal, or otherwise, that could be construed as influencing the results or the conclusions presented in this paper.

Generative AI statement

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

References

- [1] Kim HJ, Kim M-H, Park SJ, et al. (2024). Autoimmune adverse events after COVID-19 vaccination: a nationwide population-based cohort study in Korea. *J Allergy Clin Immunol.*, 153(6):1711-1720. DOI: 10.1016/j.jaci.2024.01.025
- [2] 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
- [3] Terao I, (2025) Immortal Time Bias, Confounding by Indication, and Antidepressant Pooling. *Acta Psychiatr. Scand.*, 152(4):318-319. DOI: 10.1111/acps.13827.
- [4] Akl J, Lee S, Ju HJ, et al. (2024). Estimating the burden of vitiligo: a systematic review and modelling study. *Lancet Public Health*, 9(6):e386-e396. DOI: 10.1016/S2468-2667(24)00026-4.
- [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] Lee H, Lee MH, Lee DY, et al. (2015). Prevalence of Vitiligo and Associated Comorbidities in Korea. *Yonsei Med J.*, 56(3):719-725. DOI: 10.3349/ymj.2015.56.3.719.
- [7] Sun X, Luo S, Lou L, et al. (2021). Health seeking behavior and associated factors among individuals with cough in Yiwu, China: a population-based study. *BMC Public Health*, 21:1157. DOI: 10.1186/s12889-021-11250-
- [8] Chemaitelly H, Ayoub H, Coyle P, et al. (2025) Assessing healthy vaccinee effect in COVID-19 vaccine effectiveness studies: a national cohort study in Qatar. *eLife* 2025, 14:e103690. DOI: 10.7554/eLife.103690.
- [9] Roccetti M, Cacciapuoti G,. (2025) Beyond the Gold Standard: Linear Regression and Poisson GLM Yield Identical Mortality Trends and Deaths Counts for COVID-19 in Italy: 2021–2025. *Computation*, 13(10): 233. DOI: 10.3390/computation13100233.

[10] Jones DS, Podolsky SH, (2015) The history and fate of the gold standard. *The Lancet*, 385(9977):1502-1503.
DOI: 10.1016/S0140-6736(15)60742-5.

Scope Statement:

We provide a contribution to the field of post-marketing pharmacovigilance.

Demonstration of Clinical Urgency: Our study stems from the need to counteract the threat to clinical integrity posed by a flawed finding indicating major risk for Vitiligo after Covid-19 vaccination. This spurious signal is misleading for patients, clinicians, and health authorities. By demonstrating the complete collapse of the risk signal after correct recalculation, we provide **evidence neutralizing an artifactual safety concern**.

Bias Decomposition: We introduce a Stratified Ponderation Analysis that moves beyond simple confounding acknowledgement. We decompose the initial Hazard Ratio (HR) into two necessary components: the Structural HR which is the risk attributable to demographic imbalance, and the Residual HR which serves as a direct measure of all uncorrected design failures.

Proof of Denominator Invalidity: Our investigation quantifies that the incidence rate in the reference cohort (Non-Vaccinated) is approximately 70% lower than the rate expected for its age structure. This provides **evidence of failure in the control cohort**, constituting a critical violation of common STROBE principles.

Correction of Spurious Risk: We demonstrate, through recalculation of risk, that the original, amplified HR collapses to a clinically irrelevant value. This outcome has a direct and urgent impact on **vaccine safety verification**.