Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.082258

ARTICLE

From Local Large-Scale Health Signal Inflation to Stochastic Stationarity: A Multiple-Component Risk Recalibration Framework via Intelligent Difference-in-Differences Decomposition

Marco Roccetti*

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

*Corresponding Author: Marco Roccetti. Email: marco.roccetti@unibo.it

Received: 12 March 2026; Accepted: 05 May 2026; Published: 27 May 2026

ABSTRACT: Geospatial health risk signals, characterized by associations with high magnitude statistical significance, may frequently originate from circumscribed observational data streams. When these signals are fueled by massive N-size datasets, the large dimensional scale of the sample can induce a misleading interpretation of local evidence as a statistically significant risk inflation. The objective of this study is to verify whether such health risk configurations constitute geospatial structural artifacts: namely, stochastic distortions generated by the spatial information of local health repositories that, despite their massive scale, may remain fundamentally distant from broader contextual realities. To this aim, we present a mathematical framework designed to evaluate geospatial health systemic resilience by resolving local signal inflation through a structural remodelling procedure. By integrating a Recentered Influence Function (RIF) regression and a state-tuple formalization, the model expands the analytical context beyond the boundaries of the local health data, systematically hooking into broader reference frameworks (e.g., historical metrics and national-level standards). Through the application of a spatial Difference-in-Differences (DiD) decomposition, the local risk signal is partitioned into an explained component, justified by local covariates, and a structural unexplained component. This decomposition isolates the divergence between the observed phenomena and a broader baseline, revealing how the magnitude of a local health alarm can often be a function of a measurement bias internal to the data-generating process. To prove the framework's efficacy, we conducted our analysis on a large geospatial health dataset. The analysis revealed that the initial reported alarm was the byproduct of a 32.5% structural weight deficit in the high-risk stratum of the local experimental population and, most critically, a 45.1% deficit in disease occurrences within the non-exposed local baseline compared to the national reference. This exemplar modeling has demonstrated that our proposed diagnostic and corrective framework can be a useful diagnostic tool to validate the identification of a health risk, having quantified the inversion of the original signal from an initial risk factor of 1.27 to a recalibrated 0.77 value. By isolating the structural difference between local observations and extra-local references, this methodology ensures consistency between verifiable health reality and dataset-specific outcomes, detecting and mitigating structurally inflated risk signals.

KEYWORDS: Structural health re-modelling; recentered influence function; geospatial systemic resilience; health risk recalibration; spatial difference-in-differences; intelligent signal mitigation

1 Introduction

The primary challenge in modern large-scale health modeling is the geospatial structural instability that emerges when massive administrative repositories are analyzed in isolation from their broader contextual health baselines. In contemporary high-velocity data environments, the unprecedented accumulation of

localized health data (N-size datasets) has fostered a prolific emergence of local health risk signals. However, these signals frequently originate from geospatially restricted observational streams.

When these signals are fueled by massive N-size datasets, the large dimensional scale of the sample induces a misleading interpretation of local evidence as a statistically significant risk inflation. In this context, our study posits that such configurations often constitute geospatial stochastic distortions generated by the spatial information of the original local repositories. Despite their volume, these datasets are frequently not linked with the more general, stable surrounding realities, rendering the signal a byproduct of a geospatial issue rather than a structural health reality [1].

In essence, in many of those high-throughput health architectures, the localized nature of data collection creates a condition where the outcomes become algorithmically and statistically confined to the internal correlations of a given specific geo-space. While a system may report rapid fluctuations in state determinants within a specific and local geospatial health domain, subsequent structural analyses often reveal that these groups are fundamentally decoupled from the contextual (e.g., national) baseline. This phenomenon, which we define as a geospatial non-stationarity of the sample health data, referring to the progressive loss of representativeness as a dataset becomes geographically isolated from the broader population's reality, occurs when the monitoring architecture of a local repository fails to maintain stationarity with the broader contextual ground truth [2].

The resulting non-stationary signal is thus a direct function of an internal measurement bias. In such regimes, the magnitude of the alarm is not an indicator of an observed effect size, but a measure of the mathematical distance from the local baseline. Traditional statistical power, regardless of the involved volumes, does not mitigate this decoupling; it merely provides a higher resolution for the observation of a structural divergence [3].

To address this challenge, we propose a formal mathematical framework designed to evaluate the geospatial resilience of health monitoring systems. By moving beyond traditional inferential metrics, we propose a multi-stage approach based on state-tuple formalization, Difference-in-Differences (DiD) asymmetry testing, and Blinder-Oaxaca structural decomposition [4,5]. This methodology is specifically engineered to isolate the explained component of a health signal from the unexplained geospatial structural noise. By integrating a Recentered Influence Function (RIF) projection, our model expands the analytical context, systematically comparing local health parameters with historical contextual-level reference standards [6]. This allows for a rigorous reconsideration of geospatial parameters, ensuring that the evidence is validated against verifiable, broader benchmarks rather than against repository-specific insulations. Not only that, but our multi-stage approach transforms the analysis from a qualitative observation into a quantitative measure of health signal stability. The core objective of our methodology is to quantify the structural tension between localized observations and global benchmarks, establishing whether a reported risk possesses the necessary geospatial generalizability to be considered a stable finding. Furthermore, our framework is also designed to determine the model recalibration feasibility of a health signal: identifying whether a divergence from the context can be corrected through recalibration or represents a critical systemic divergence within the monitoring architecture.

To demonstrate the efficacy of this approach, we applied the framework to a case study that is uniquely significant due to its massive scale (approximately 3 million individuals) situated in the Asia-Pacific region. This case is of particular interest as it utilizes a large local repository to propose associations between a specific determinant (vaccination) and adverse effects (cancer), providing the ideal high-stress environment for our audit. By isolating this structural gap between the local observational frequency and the extra-local incidence mean, our methodology has successfully neutralized, in this specific case, the architectural divergences of the local health repository. This led to individuate a structural flip, where the initial local risk signal reversed

direction upon recalibration, proving the finding was a geospatially limited signal rather than a generalizable property. Ultimately, this massive dataset served as an optimal exemplar to demonstrate that our framework can effectively identify when the local baseline is the result of an unbalanced data extraction, ensuring that any localized risk assessment is in relationship with the reality of a broader (e.g., historical and national) reality. In particular, in our scrutinized case, our investigation revealed that the initial reported alarm was the byproduct of a 32.5% structural weight deficit in the elderly stratum of the experimental population and, most critically, a 45.1% deficit in the number of disease occurrences within the non-exposed portion of that portion of the local population compared to the national standard.

Beyond this exemplar, large-N localized biases are prevalent across various fields. Literature confirms that without formal analysis of structural variance, massive databases risk generating non-stable correlations [7]. Consequently, rigorous analytical frameworks are essential to ensure model stability and to validate data-driven findings. Structural decompositions have previously identified geospatial gaps in socio-medical fields, such as obesity disparity [8] and child nutrition [9], where unexplained components were artifacts of the data extraction procedures. In oncology, researchers have likewise warned that high-dimensional volumes do not guarantee veracity [10]. Therefore, we reiterate that RIF-based recalibration and structural decompositions can act as conditions for estimand identifiability to validate any large-scale, yet local, risk assessment.

It is also important to mention that this paper addresses the problem of geospatial signal instability through a systems engineering approach. Unlike traditional health-related inquiries, this study deliberately adopts systems engineering and structural modeling perspectives to investigate the integrity of health risk signals. By employing a technical terminology and a mathematical framework consistent with structural analysis, we aim to provide a formal explanation of the systemic mechanisms that govern the relationship between local data streams and their broader contextual references. This approach is essential to quantify the resilience of information systems against the emergence of geospatial issues and ensures that the adopted structural terminology is technically required to explain the complex informational dynamics described in the following sections.

Before concluding this section, it is finally worth mentioning that the primary methodological novelty of this work lies in the hierarchical orchestration of traditionally disparate structural tools into a unified mathematical and diagnostic pipeline. By transitioning from a simple discovery mindset to a structural auditing perspective, the proposed framework provides a formal mechanism to quantify the resilience of health risk signals when projected across different geospatial scales. This approach allows for the identification of systemic distortions that remain invisible to standard inferential techniques.

The remainder of this paper is as follows. In Section 2, we present our methodology with a maximum of equilibrium between mathematical accuracy and the comprehension of sophisticated concepts. In Section 3 we demonstrate the efficacy of our approach with a very exemplary case. In Section 4, we provide a discussion summarizing all the relevant issues of our study. Finally, Section 5 terminates the paper.

2 Methods: Geo-Spatial Structural Stability Analysis

The fundamental challenge in large-scale administrative health modeling is the potential for the emergence of local health risk signals, characterized by associations with statistical significance, which frequently originate from circumscribed observational data streams. When these signals are fueled by massive N-size datasets, the dimensional scale of the sample may induce an interpretation of local evidence as a high-magnitude risk inflation. Unfortunately, these risk configurations often simply constitute stochastic distortions generated by the geospatial information of the original local repositories that, despite their mass, are unanchored from the general surrounding realities.

In more precise words, we operate in a scenario where a given specific determinant *(D)*, such as a massive medical intervention or a localized environmental exposure, acts upon a defined portion of a local (*L*) experimental population. This interaction generates a vast amount of localized data and observations, from which a local risk factor (*RF_OBS*) is computed.

When this *RF_OBS* exceeds a given threshold (take unity as an example), the system emits an automatic danger warning, suggesting an association between the determinant and a negative health outcome. However, the large volume of local data does not automatically guarantee the structural stability of that signal. The challenge lies in understanding not only whether this potential local danger signal can be extended to a broader geospatial context, such as a national one, but also how it reacts structurally when the local information upon which it is based is replaced by broader contextual references that would, in fact, render it generalizable. In essence, in complex geospatial health systems, a large-scale local signal can be an endogenous feedback of sampling biases rather than a broadly extendible health reality.

In this context, the objective of our methodology is to determine whether this kind of warning is a robust property of the determinant *D* or a byproduct of the system's loss of integrity. To do this, we project the local components of *L* onto a global context (*C*). We analyze whether the signal remains stable when the local experimental population is re-embedded within a broader, more stable (e.g., national) baseline. If the signal deviates during this projection (either evaporating or exploding), we should conclude that the original *RF_OBS* must not be interpreted as the outcome of a stable signal, but as a limited geospatial issue progressively losing representativeness when the observation scales to a broader context.

To analyze this stability, we treat the health system through a dynamic state characterized by a multi-dimensional tuple. Elements of this tuple act as follows. *D* is the determinant, that is the external factor introduced into the health system *S*, for example a massive COVID-19 vaccination, which is extended all over the local portion of an experimental population of size *N*. *OBS_L* are the local observations of the occurrence of a given disease (e.g., cancer cases) per 10,000 individuals following *D*. Those detected effects (or cases), can be structured, in turn, into a dual-level hierarchy based on groups (Exposed *E* and Non-Exposed *NE* and strata, Young (*young* < 65) and Elderly (*older* ≥ 65). With the *w* parameters, instead, we indicate the portions of elderly in the local experimental population compared to the national baseline, that is: *w_*(*L*, *older*) and *w_*(*C*, *older*).

RF_OBS is in essence the computed risk factor based on local data and relative observations, that is the internal metric of the local system's response, while *RRF* is the recalibrated risk factor, in simple terms the corrected state after recalibrating local observations based on the contextual references provided by *C*, where the occurrence of the very same disease is historically registered and known as *I_C*. All this allows us to move beyond simple statistics into the field of integrity systemic modeling. The tuple could, hence, take the following form:

$$S =< L, C, D, OBS, OBS\_E, OBS\_NE, N, I\_C, RF\_OBS, RRF > \quad (1)$$

Notable in this framework are the following facts: *L* vs. *C* represents the tension between the local experimental domain and the national gold-standard stability. *RF_OBS* is the initial danger warning derived strictly only from local data and observations. *RRF* (the Recalibrated Risk Factor) represents the final state of the signal after being filtered through the contextual weights/references provided in the broader context of *C*. Obviously, any discrepancy between *RF_OBS* and *RRF* is a sign of a structural non-integrity state. The goal of our mathematical operators, described below as an integral part of our methodology, is to measure the depth of this discrepancy.

To be considered is also that the proposed methodology operates under a stationarity assumption, which posits that the global benchmark (*C*) represents a stable, historical equilibrium of a given disease's natural history. Therefore, the validity of the structural recalibration relies on the assumption that the global baseline serves as a constant and reliable counterfactual for any localized experimental setting. Within this paradigm, the analytical priority shifts from measuring stochastic uncertainty to evaluating the structural resilience of the signal. In this sense, it is worth reiterating that, in terms of formal estimation, our framework does not seek to establish a causal link between the determinant *D* and the outcome. Instead, it just estimates a structural bias, conceived as the residual variance that remains when a local risk signal is projected onto a contextual baseline. The core assumption of our model, in fact, is the contextual stationarity, assuming the national benchmark can serve as a stable counterfactual against which the integrity of the local data-generating process is tested.

To begin with the narrative of our methodology, a preliminary check to which the local signal should be subjected is a classical representativeness test drawn by classical inferential statistics and concerning the Non-Exposed (*NE*) arm of the experimental population. In any valid health risk assessment, in fact, the *NE* group must act as a reliable counterfactual, that is, a kind of mirror of the population's natural history. If the local system fails to reflect the *NE* group accurately, the denominator at the basis of the classical calculation of a risk factor (i.e., number of occurrences of a disease within *E* over the same figure for *NE*) becomes a relevant preliminary index that warns about the structural stability of the investigated signal. We quantify this using the well-known baseline divergence operator (*Z*) from inferential statistics [11]:

$$Z = \frac{OBS\_NE - I\_C}{\sqrt{(1 - I\_C) \times I\_C / N}}. \tag{2}$$

where *OBS_NE* represents the observed incidence in the local non-exposed group, *I_C* is the national incidence (contextual baseline), and *N* is the sample size of the local repository.

From traditional inference statistics, we know that if $|Z| > 1.96$, the local non-exposed group should be interpreted as a structural outlier, with the meaning of this first check being as follows. The initial danger warning represented by *RF_OBS* has probably been pushed up, not by an increase in cases among the *E* portion of the experimental population, but by a kind of structural baseline depletion in the *NE* group, leading to a signal for which instability is highly predictable.

If such a divergence is detected, we must then determine if this observation instability is either symmetric or asymmetric, that is confined only to the local context of our investigated health geospace. To this aim, we decided to use the structural Difference-in-Differences (*DiD*) operator, leveraging on well known data analysis methodologies, to measure the relative deviation of the local components from a broader (e.g., national) context, using the following couple of subsequent equations:

$$Diff\_L = (OBS\_L - I\_C) / I\_C \tag{3}$$

$$DiD = Diff\_(L, E) - Diff\_(L, NE) \tag{4}$$

here the meaning of the *Diff_L* operator is that of measuring the relative deviation between local observations (*OBS_L*) and the contextual benchmark (*I_C*). Instead, the *DiD* operator, by subtracting the deficit of the Non-Exposed group (*NE*) from that of the Exposed group (*E*), isolates the divergence in monitoring density between the two cohorts. Again, the logic is simple: if *DiD* is different from zero, then the observation mechanism is asymmetrical. Asymmetry, in this case, is another warning about the instability of the original data as it corresponds to a situation where one arm of the dataset (typically the *NE*) is significantly less monitored than the other. This creates a mechanical bias where the relative *RF_OBS* is not measuring a stable

health system condition; indeed, it is measuring the progressive divergence between L and C that emerges as soon as contextual observations from C are integrated into the monitoring system.

At this point, to further understand the parametric decomposition of the original risk signal, we can subsequently refer to the well-known Blinder-Oaxaca signal decomposition applied to decompose *RF_OBS* and see what proportion of the risk is justified by the population' profiles vs. how much is a byproduct of the structural noise within the local dataset. To analyze how *RF_OBS* reacts structurally when projected onto contextual benchmarks, we linearize the signal through a logarithmic transformation. In this framework, we decompose the total gap observed in the local health system into two distinct vectors: one driven by the compositional weights (termed w, see definitions above) of the local groups and one driven by the structural parameters (b) of the observation mechanism. Using the components defined in the state tuple S above, the decomposition of the observed risk can be formalized as follows.

$$\ln(RF\_OBS) = \sum w\_(L,E)(b\_(L,E) - b\_(L,NE)) + \sum b\_(L,NE)(w\_(L,E) - w\_(L,NE)) \tag{5}$$

where: $w\_(L, E)$ and $w\_(L, NE)$ represent the normalized distributional weights (e.g., the age-stratified proportions) of the Exposed and Non-Exposed groups within the local domain L, while $b\_(L, E)$ and $b\_(L, NE)$ represent the structural coefficients estimated by the model, capturing the rate of the disease case occurrence for each unit of weight.

The systemic interpretation of this stability control is that $\sum b\_(L,NE)(w\_(L,E) - w\_(L,E))$ is the term that quantifies the portion of the *RF_OBS* signal that is justified by the actual difference in the demographic mass between the two arms. If the signal is primarily driven by this component, it suggests that the risk is a distributional property of the local experimental population. Instead, $\sum w\_(L,E)(b\_(L,E) - b\_(L,NE))$ is the portion that accounts for the critical noise of the system. It measures how the same unit of population weight (w) produces divergent risk estimates (b) due to the decoupled state of the local baseline with respect to the broader (national) context.

In the end, the core challenge of this check is to understand how the local result *RF_OBS* reacts structurally when local parameters are replaced by broader contextual benchmarks. By substituting the local weights w_L (drawn from L) with the national reference weights w_C (for example, the national standards), the real generalizability of the original *RF* signal is tested.

As a preliminary step, we can operationalize this passage to adapt it to the context of the exemplar case we will run in the next section, by providing the following formula to use which becomes as follows:

$$Stability\ Check = OBS\_L / I\_C \tag{6}$$

Note that in this Eq. (6) the denominator has been simplified for clarity. Replacing the local denominator with the national standard reveals if the signal survives in a broad context. In fact, if the new resulting signal decreases below unity during this substitution, this implies that we have reached the so-called *structural flip*, which proves that the original *RF_OBS* was not a property of the determinant D (the exposure), but a byproduct of the specific observation gaps within the local data, regardless of its relative size N. In this case, the initial danger warning should be confined to a simple local geospatial result, as it lacks the structural support to survive the projection into a broader (national) context.

The final stage of our methodology is the recalibration using the aforementioned weights w_C. We use the recentered influence function (*RIF*) to project the local observations (OBS_L) onto the national contextual weights (w_C). This acts, in some sense, as the inverse of the previous formula, with the intent of attempting to repair/reconstruct a more stable *RRF*. The corresponding formula is as follows:

$$RRF = \sum OBS\_L \times w\_C \tag{7}$$

The resulting value represents the mathematical intersection between a local observation (*OBS_L*) and the global stability (*w_C*). This product is a filter designed to test the robustness of the individuated anomaly: it ensures that a local signal is considered valid only if the injection of national weights produces a reinforced product capable of reaching the system's ground truth. If the recalibrated product fails to reach the national benchmark, despite this systemic amplification, the local incidence is immediately exposed as a geospatial instability. An insufficient product zeros out the initial risk, proving that the national support is inadequate to validate the local peak. Failing to minimize the *L*/*C* structural variance definitely identifies the risk as a loss of systemic integrity.

As an important note, it is essential to highlight that, since in complex geospatial health systems, a large-scale local signal can be a feedback of internal sampling biases rather than a broadly generalizable reality, the choice of integrating *DiD*, *RIF* regression, and Blinder-Oaxaca decomposition is driven by the necessity to observe this multifaceted reality from different perspectives. While *DiD* isolates the asymmetry in observation density between cohorts, *RIF* is essential for the contextual recalibration of local incidences. This specific combination provides the mathematical bridge required to project non-stationary local results onto stable national benchmarks, a step that is technically unattainable through the isolated use of these methodologies alone.

In closing, to facilitate a full comprehension of our methodology, we summarize in Table 1 its workflow with all the main steps we have proposed.

Table 1: Four-stage diagnostic pipeline for geospatial structural integrity auditing.

Stage	Operator/Objective/Metric
1: Baseline Analysis	Diagnostic Operator: Baseline Divergence Test
	Operational Objective: Detect if the local non-exposed group (*NE*) is a structural outlier relative to the national context.
	Key Metric/Equation: Z-score Eq. (2)
2: Asymmetry Check	Diagnostic Operator: Structural *DiD*
	Operational Objective: Measure discrepancies in monitoring density and case-capture between cohorts (*E* vs. NE)
	Key Metric/Equation: *DiD* Gap Eqs. (3) and (4)
3: Signal Anatomy	Diagnostic Operator: Blinder-Oaxaca Decomposition
	Operational Objective: Isolate compositional effects (population profile) from structural noise inherent to the local dataset
	Key Metric/Equation: Residual Vector Eq. (5)
4: Integrity Repair	Diagnostic Operator: *RIF* Contextual Projection
	Operational Objective: Test signal resilience (structural flip) and recalibrate local observations against stable national benchmarks
	Key Metric/Equation: *RRF* Eqs. (6) and (7)

3 Experiments, Computational Results and Source Data

To evaluate the proposed framework in the field, we tested the health system state S using a large-scale repository (N approx. equal to 3×10^6 individuals). In particular, we instantiated the L, C contexts and the D determinant, with all their relative values, in terms of local and national disease incidences, and original data from the following references [12–16] and relative datasets, as exposed in the Availability of Data and Materials Section where a health dataset (the National Health Insurance Service NHIS) of a major city (Seoul) in the Asia-Pacific region (South Korea) is referenced. It is also worthwhile noticing that the determinant D represented a COVID-19 vaccination, and the observations of cases and relative incidences pertained to serious adverse effects, specifically, cancer.

The first step of our analysis has been to compare the local observed incidences OBS_L and the demographic weights w_L against the national contextual benchmarks I_C, as well as gold-standard weights w_C, using the data summarized in Table 2 below.

Table 2: Health system instantiation and structural data baseline.

System Component	Local OBS_L (Incidence/10,000)	Contextual I_C (Expected/10,000)	Deficit ($Diff\_L$)	Local Weight w_L	Context Weight w_C
High-Risk Stratum (*Older*, *NE*)	85.2	155.2	−45.1%	12.15%	18%
High-Risk Stratum (*Older*, *E*)	113.5	155.2	−26.9%	12.15%	18%
System Volume (*Elderly*)	12.15%	18%	−32.5[1]%	100%	100%

Note: [1] Achieved as (12.15 − 18)/18 = −32.5.

The following structural engineering interpretations of the results of Table 1 are in order. The health system locally suffers from a 32.5% volume deficit in its most critical observation stratum (the elderly). Not only that, but the *NE* arm has a deficit (45.1%) in the elderly, which is structurally inflated with respect to the other arm *E* (26.9%). This generally indicates that the dataset is structurally anemic, even before any further analysis begins. We now pass to the preliminary local signal analysis suggested by Eq. (2) to test the local state divergence from the context. We begin by reminding that the initial local risk factor was computed as $RF\_OBS = 1.27$ (42.62/33.43). To this aim, we tested the integrity of the *NE* baseline through Eq. (2), as shown in Table 3.

Table 3: Preliminary signal and local state divergence—Eq. (2).

Variable	Value	Calculation	Statistical Significance
Local *NE* Rate	0.003343	33.43/10,000	N/A
Global I_C Rate	0.005502	55.02/10,000	N/A
Z-Score Eq. (2)	−225.1	$\frac{0.003343 - 0.005502}{\sqrt{(1 - 0.005502) \times 0.005502/595,007}}$	$p < 0.0001$

In this case, the structural engineering interpretation is that the extreme value of −225.1 provides evidence to the fact that the *NE* group is an outlier and the initial danger warning is a candidate to be a mathematical byproduct of a baseline that is 40% lower than the ground truth.

We now pass to the use of Eqs. (3) and (4) to check if the health system locally monitors *E* and *NE* with the same precision, whose results are reported in Table 4.

Table 4: Asymmetry and mechanical bias—Eqs. (3) and (4).

Parameter	Calculation	Result	Structural Equation
*Diff*_(*L*, *E*)	(113.5 − 155.2)/155.2	−26.9%	Eq. (3)
*Diff*_(*L*, *NE*)	(85.2 − 155.2)/155.2	−45.1%	Eq. (3)
DiD Gap	(−26.9%) − (−45.1%)	+18.2%	Eq. (4)

Here, the most relevant engineering interpretation is the positive 18.2% *DiD* gap (being the other relative deficits known since before). This gap confirms an asymmetric monitoring instability. The health system locally captures cases in the *E* group much better than in the *NE* group, probably creating a mechanical inflation of the risk factor.

We now shift to the signal decomposition as per Eq. (5). In essence, we analyze the internal decomposition of the initial 1.27 signal risk to see if it is justified by the population profiles (i.e., the *w* components). We should remind that the *w* coefficients represent the mass of the elderly in the local dataset (12.15%), while the *b* coefficient represent the case incidences for that specific group (computed per 10,000 individuals). Thus, to make a decomposition of the *RF_OBS* = 1.27, the coefficients *w* and *b* have to be computed as average values for the critical stratum (i.e., elderly, ≥65 years). In particular, we should remind that: *w_L* = 0.1215 and *w_C* = 0.1800. Moreover, *b*_(*L*, *E*) = 113.5 and *b*_(*C*, *E*) = 155.2, while *b*_(*L*, *NE*) = 85.2 and *b*_(*C*, *NE*) = 155.2. With these starting points, the results from applications of Eq. (5) are those shown in Table 5.

Table 5: Detailed risk signal decomposition calculation—(Eq. (5)).

Vector Component	Mathematical Expansion	Numerical Result	Relevance
Compositional (*w*)	$b\_(L, NE) \times (w\_(L, E) - w\_(L, NE)) = 85.2 \times (0.1215 - 0.1215)$	0.0	Zero impact from internal balancing
Structural Noise (*b*)	$w\_(L, E) \times (b\_(L, E) - b\_(L, NE)) = 0.1215 \times (113.5 - 85.2)$	3.43	Risk is a pure structural gap
Total Residual	$(b\_(L, E) - b\_(L, NE))/b\_(L, NE) = (113.5 - 85.2)/85.2$	0.33	+33% relative error in high-risk stratum

The takeaway here is that the parametric decomposition of the risk signal shows that the risk engine is not the demographic distribution (the *w* term is zero due to the PSM matching applied to the original dataset [17]), whereas the structural divergence *b* between the two monitoring arms drives the artificial inflation of the observed risk.

Essentially, the health system, with its computed risk, is only measuring the differential efficiency of the sensors (*E* vs. *NE*), rather than the effect of the determinant *D* over the experimental population as a whole.

It is now the turn to test the generalizability of the local initial risk signal by substituting the local deflated *NE* arm with the national contextual anchor (*I_C* = 55.02), as per Eq. (6), thus achieving the results portrayed in Table 6.

Table 6: Stability test—Eq. (6).

Test Condition	Mathematical Operation	Result	Systemic Status
Local Perception *RF*_(*OBS*)	42.62/33.43	1.27	Alarm Warning
Contextual Reality Recomputed *RF*	42.62/55.02	0.77	Reduction of the alarm

The engineering interpretation is that when projected onto the global context, the initial signal decreases from a 27% increase to a 23% deficit. This transition identifies *RF*_(*OBS*) as a geospatial local value lacking the structural support to survive the projection into a broader national context.

We can now enter the final phase of our methodology and attempt to repair the local incidence by injecting the national contextual weights (w_C = 18%) and correcting for the detected asymmetry (*DiD* = 18.2%) following Eq. (7), and yielding the results shown in Table 7.

Table 7: RIF recalibration: local observations vs. contextual weights—Eq. (7).

Variable	Symbol	Numerical Value	Calculation/Reference
Local Base	*OBS_L*	31.2	Local cases per 10,000 (Younger + Older)
Contextual Benchmark	*I_C*	55.02	Contextual Incidence
Contextual Weight	*w*_(*C, older*)	0.18	National Standard (18% Elderly)
Asymmetry Factor	$1 + \Delta b$	1.182	Derived from *DiD* (+18.2%)
Final *RRF*	*RRF*	36.88	31.2 × 1.182

Here, the operational meaning is that the attempt to repair the system by re-weighting the *NE* observations to match the 18.00% weight of the national *C*, using our *RRF* operator and considering that: for the entire (younger + older) experimental population *OBS*_(*L*, *NE*) = 31.2 (per 10,000), *I_C* = 55.02 (per 10,000) and *w*_(*C*, *older*) = 18%, yielded a *RRF* = (31.2 × (1 + 0.182)) = 36.88.

In synthesis, this recalibrated value remains profoundly decoupled from the contextual ground truth (36.88 vs. 55.02). This calculation has definitely shown that even after mathematically correcting the asymmetry and adjusting for the missing elderly mass, the initial local results have remained very far from the broader national target. This further confirms that the selected subjects in the local dataset were qualitatively non-representative of the health state of the general population, resulting in a total loss of systemic health integrity. This can be seen as the final evidence that, in our exemplar case, the proposed methodology succeeded in putting in evidence the deviation between the local and the global geospatial contexts. We conclude this section by eliminating any possible ambiguity regarding the origin of the employed parameters and relative information, presenting an extended tabulation of all the data which are reported in Table 8 as drawn from the already cited references [12–16] and relative datasets referenced in in the Availability of Data and Materials Section at the end of this manuscript.

As evidenced by the data in Table 8, all primary information, including participant counts, disease occurrences, and stratified incidence rates, is provided there in an extended form. This ensures that the structural flip and the subsequent recalibration are fully verifiable and replicable, grounded in the transparent reporting of the local and national parameters used within our diagnostic framework.

Table 8: Reference data.

Category	Global (*C*)	Local *NE*	Local *E*
Elderly ($w \geq 65$)	18.00%	12.15%	12.15%
Young ($w < 65$)	82.00%	87.85%	87.85%
Participants (Young)	N/A	522,722	2,090,888
Participants (Elderly)	N/A	72,285	289,140
Total Participants	N/A	595,007	2,380,028
Cases (Young)	N/A	1373	6861
Cases (Elderly)	N/A	616	3283
Total Cases	N/A	1989	10,144
Incidence Rate (Young)	33.03/10,000	26.27/10,000	32.81/10,000
Incidence Rate (Elderly)	155.20/10,000	85.22/10,000	113.54/10,000
Total Incidence Rate	55.02/10,000	33.43/10,000	42.62/10,000

4 Discussion

The primary objective of this study was to evaluate the systemic resilience of health monitoring frameworks when large-scale local datasets are projected onto global contextual baselines. In complex geospatial modeling, the reliability of a signal is not a function of its statistical significance alone, but of its structural stability across different scales of observation. Our approach addresses this problem by implementing a multi-stage audit framework based on state-tuple formalization, Difference-in-Differences (*DiD*) asymmetry testing, and Blinder-Oaxaca structural decomposition. This methodology is designed to identify possible spatial non-stationarity in data integrity, where localized observation biases may generate high-magnitude artifacts that fail to survive contextual recalibration.

Only to rigorously test the effectiveness of this framework, we selected a case study characterized by a very massive dimensional scale, providing a sufficiently stressed environment to demonstrate how structural instability can override nominal statistical power. The structural analysis conducted in our study led to three conclusions regarding the integrity of that large-scale administrative dataset, regardless of its massive size.

The first of those results is relative to the health systemic decoupling (local/global) and relative inferential integrity. In particular, our result revealed that the 32.5% volume deficit in the high-risk stratum of the local experimental population has created a decoupling of the local system (L) from the global context (C). By under-representing the elderly population by nearly a third, the local baseline was fundamentally shifted. What it is of general interest, beyond the example, is that, while the original dataset may show high reporting compliance, its inferential integrity, defined as its capacity to support generalizable conclusions, may decrease when generalizing from an unanchored, albeit enormous, starting dataset to broader contexts.

Nonetheless, in the example we ran, the most critical finding remains the 18.2% *DiD* gap. We have provided evidence that the observation mechanism in the *NE* arm was deflated by 45.1%, relative to the national context, while the *E* group showed a smaller negative variation (26.9%). This asymmetry was evidence of the fact that the initial alarming risk signal ($RF\_OBS = 1.27$) was a byproduct of the *NE* group's inability to capture cases (cancers). Since, it is structurally unsound to compare an unreliable non-exposed group with a less unreliable exposed group, the classification of the initial signal as a structural deviation was confirmed by this identifiable 18.2% *DiD* gap, derived via Eq. (4), see Table 4. This value in fact makes explicit that the observed risk was not a biological property but a byproduct of the monitoring system's inability to maintain a symmetric baseline between *E* and *NE*.

Another important result we achieved was the usefulness of the application of a Blinder-Oaxaca approach to understand the distinction between compositional effects (what the model balances through internal algorithms) and structural effects (what the local dataset inherently represents). In our case, the fact that the entire risk gap was captured by the residual structural vector, identifiable via the decomposition in Eq. (5), see Table 5, has brought evidence to the hypothesis that the original risk signal was not driven by the variables under algorithmic control, but by a structural divergence (total residual = 0.33) that no amount of internal balancing could mitigate.

In this sense, the decomposition has acted well by revealing that the signal was of a purely computational nature, thus confirming a general rule according to which excessive technical sophistication in internal algorithmic loops can (occasionally) mask foundational descriptive divergences from reality.

Finally, of relevance was also our original idea to resort to a recentered influence function (*RIF*) projection to understand if a compromised health system can be effectively repaired. Our case has demonstrated that even when we mathematically correct the demographic local weights to match the national standard (18%), the resulting incidence can remain much lower (33%) than the national ground truth (36.88 vs. 55.02).

This suggests that our technique is able to capture situations where the problem is not merely a quantitative lack of data, but a qualitative loss of integrity in the local selection process. In our scrutinized case, if a risk signal is so baseline-dependent that it triggers a structural flip, re-weighted from 1.27 down to 0.77 as per the RIF operator in Eq. (6), upon contextual projection, the results must be considered as confined to a selected, non-representative sub-sample, because the signal lacks the resilience required for generalizability. In these situations, observed structural flips down to a sub-threshold of 0.77 may carry relevant implications for healthcare policy. This kind of shift would suggest, in fact, that local findings, even if statistically significant, may be weak signals losing their robustness when calibrated against national benchmarks. For policymakers, this means that local spikes in risk should not trigger immediate, large-scale policy changes. A weak signal identifies only a potential area of concern that requires further empirical confirmation through repeated experiments or multi-center validation. Only when that signal maintains its strength across both local and national contexts should it be considered a strong driver for systemic policy interventions, thereby preventing the misallocation of resources based on localized non-stationary alarms.

To conclude this section, we acknowledge the boundaries of this study. We reiterate our complete agnosticism regarding the specific determinant D, as our analysis demonstrates the limits of dataset stability rather than biological factors. However, it is important to emphasize that our methodology is designed to be domain-agnostic; its diagnostic power lies in the ability to stay silent in stable systems and only trigger a recalibration when a structural decoupling is detected, regardless of the specific geospatial field of application. In this sense, the case analyzed has only served as a single, albeit massive, exemplary instance. Furthermore, it is important to acknowledge that the reliability of the national benchmarks has been assumed as the standard reference for our recalibration. We are aware that while national registries are subject to rigorous validation protocols, they are not immune to reporting lags or various systematic biases. However, in our context of geospatial structural analysis, these benchmarks still represent the most stable reference of the health system's state. Even if not perfect, they serve as the necessary contextual baseline to identify non-stationary deviations in localized datasets.

Insisting on limitations, it is worth emphasizing that our specific application of the proposed methodology has operated at the aggregate level using public national standards. While our findings highlighted a relevant divergence in the experimental design of a given case, a granular inspection would require access to individual-level records, which remains a further recommendation for future independent research. On a more numerical side, instead, in regimes characterized by extreme statistical deficits (e.g., $Z = -225.1$), traditional measures of uncertainty like Confidence Intervals can become secondary to the structural depletion of the point estimate itself.

Finally, to summarize the evidence extracted from our case, the following Table 9 synthesizes the four core pillars that exemplify how our methodology has been able to extract the signs of structural instability from the local initial geospatial signal and its subsequent recalibration of the scrutinized case.

Table 9: Summary of findings/techniques.

Metric	Result	Engineering Significance
Baseline Deficit	−45.1%	Mathematical void in cases within the local *NE* baseline compared to context
Structural Divergence	$Z = -225.1$	Local *NE* as a terminal outlier.
Structural Flip	1.27 → 0.77	Risk signal reverses direction upon Blinder-Oaxaca projection
Structural Weight Deficit	−32.5%	Massive deflation in the high-risk (elderly) stratum volume

We conclude by positing that future research should integrate the decomposition-based analyses we have proposed herein as a standard pre-processing step for large-scale observational studies ensuring that evidence remains grounded in verifiable reality rather than in dataset-specific artifacts [18–20]. Fig. 1 provides a graphical summary of our research aim and the resulting methodological architecture. The input layer (top, left) ingests localized massive-N geospatial health signals. The core framework (top, right) performs a dual-step structural analysis via *RIF-DiD* decomposition and Blinder-Oaxaca engine adjustment to isolate stochastic distortions from the true signal. The process culminates (bottom, right) in the generation of a stationarity-corrected risk metric, thereby ensuring the reliability of the geospatial health hazard assessment.

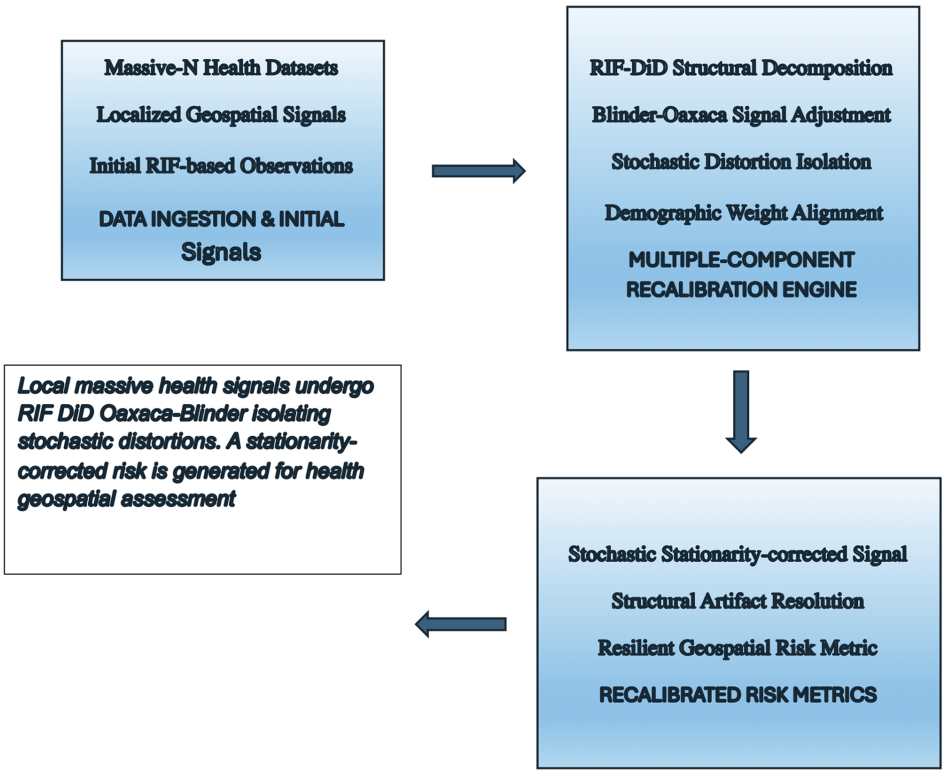

Figure 1: Architectural workflow of the multiple-component risk recalibration framework.

5 Conclusion

In this study, we have introduced a diagnostic computational framework for analyzing the structural integrity of large-scale geospatial health systems. By formalizing the health system as a multi-dimensional state tuple and implementing a sequence of diagnostic operators, including a baseline divergence test, a structural difference-in-differences evaluation, and a Blinder-Oaxaca decomposition, we have demonstrated how to detect and mitigate structurally inflated risk signals originating from geospatial instabilities. This methodology shifts the focus from simple internal statistical significance to a broader concept of data integrity, providing a rigorous auditing protocol to evaluate whether a localized data stream possesses the structural reliability required to support generalizable conclusions.

The application of this framework to a massive dataset has served as a critical stress-test, proving that even extreme dimensional scales cannot compensate for foundational decoupling from global contextual benchmarks. Our analysis successfully isolated a high-magnitude risk signal and identified it as a mathematical byproduct of asymmetric monitoring and baseline depletion. Through the use of recentered influence functions (*RIF*), we further demonstrated that structural issues in data selection are often irreparable by simple re-weighting, confirming that the integrity of a health system depends more on its qualitative consistency with contextual reality than on the quantitative volume of its local observations.

A limitation of this study is the reliance on the temporal availability of national benchmark updates, which may introduce minor reporting lags in the recalibration process. Furthermore, while the current analysis leverages a massive geospatial dataset, it serves as a proof-of-concept for the structural audit protocol, requiring further validation across diverse international jurisdictions to confirm its full generalizability. It is also important to differentiate our framework from other existing spatial methodologies. While Bayesian disease mapping and hierarchical models are effective at smoothing stochastic noise in small-area estimation, they often lack the diagnostic power to identify systemic reporting gaps in massive administrative datasets. Similarly, traditional calibration and reweighting methods (like Propensity Score Weighting) focus on internal balancing of covariates but remain blind to the structural decoupling between a local repository and its global context. Our approach, by integrating RIF-based recalibration with Blinder-Oaxaca decomposition, provides a unique diagnostic pipeline that isolates the structural flip, a phenomenon that standard spatial regressions cannot detect, as they typically treat the local baseline as a fixed, reliable reference rather than a potentially non-stationary artifact.

In conclusion, we maintain that the integration of decomposition-based analyses could become a useful diagnostic protocol for any large-scale observational studies in the geospatial health field. As administrative registries continue to grow in size and complexity, the risk of local-sample selection biases, where internal algorithmic sophistication masks structural divergence, may become increasingly prevalent. Our proposed framework offers a necessary metric for geospatial health engineering, ensuring that automated danger warnings are grounded in stable, verifiable broader realities. This approach provides a fundamental tool for researchers and policy-makers to safeguard the resilience of global health information systems against the emergence of structural non-stationarity.

Acknowledgement: None.

Funding Statement: The author received no specific funding for this study.

Availability of Data and Materials: All the data supporting the findings of this study are available from the referenced papers [12–16] and from the direct links to all the following datasets: (a) World Bank. Population ages 65 and above Republic of Korea (https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS?locations=KR); (b) Statista. Republic of Korea: Cancer crude incidence rate by age, 2022 (https://www.statista.com/statistics/1440818/

south-korea-cancer-crude-incidence-rate-by-age/) and (c) Supplementary Material III, Table S4 (https://static-content.springer.com/esm/art%3A10.1186%2Fs40364-025-00831-w/MediaObjects/40364_2025_831_MOESM3_ESM.docx).

Ethics Approval: Not applicable as this study does not involve humans, animals, plants or individual data.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript (alphabetic order):

b	Structural coefficient (rate of disease occurrence per unit of weight)
b_(L, E)	Structural disease occurrence rate per unit of population weight within the *E* group
b_(L, NE)	Structural disease occurrence rate per unit of population weight within the *NE* group
C	Context (Global/National reference)
D	Determinant (External factor/Exposure)
DiD	Difference-in-Differences
Diff	Relative deviation (Difference) operator
Diff_(L, E)	Observation gap in the *E* relative to the contextual ground truth
Diff_(L, NE)	Observation gap in the *NE* relative to the contextual ground truth
E	Exposed group
I_C	Contextual Incidence (National ground truth/Expected rate)
L	Local (Experimental domain)
N	Experimental Population size (Dimensional scale)
NE	Non-Exposed group (Counterfactual/Control arm)
OBS	Observations of disease occurrences
OBS_E	Observations in the Exposed group
OBS_L	Local Observations (Disease occurrence per 10,000)
OBS_NE	Observations in the Non-Exposed group
PSM	Propensity Score Matching
RF	Risk Factor
RF_OBS	Observed Risk Factor (Initial local danger warning)
RIF	Recentered Influence Function
RRF	Recalibrated Risk Factor (Corrected signal after contextual filtering)
S	Health System State (Formalized as a multi-dimensional tuple)
w	Distributional weight (e.g., age-stratified demographic proportions)
w_C	Contextual/National weights
w_(C, older)	Percentage of elderly individuals according to the national context
w_L	Local weights
w_(L, older)	Percentage of elderly individuals within the local dataset
w_(L, E)	Age-stratified proportions of the *E* group within the local dataset
w_(L, NE)	Age-stratified proportions of the *NE* group within the local dataset
Z	Baseline Divergence Operator (Z-score test for representativeness)

References

1. Arnout VD, der Laan Jan V, Annemarie P. Detecting reporting errors in data from decentralised autonomous administrations with an application to hospital data. J off Stat. 2018;34(4):863–88.
2. Prosperi M, Guo Y, Sperrin M, Koopman JS, Min JS, He X, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nat Mach Intell. 2020;2(7):369–75. doi:10.1038/s42256-020-0197-y.

3. Roccetti M. 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.
4. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758–64. doi:10.1093/aje/kwv254.
5. Jann B. The Blinder-Oaxaca decomposition for linear regression models. Stata J Promot Commun Stat Stata. 2008;8(4):453–79. doi:10.1177/1536867x0800800401.
6. Firpo S, Fortin NM, Lemieux T. Unconditional quantile regressions. Econometrica. 2009;8(4):436–55. doi:10.3982/ECTA6822.
7. Ye N, Li K, Wu F, Zhou J, Bai H, Yu R, et al. OoDBench+: quantifying and understanding two dimensions of out-of-distribution generalization. IEEE Trans Pattern Anal Mach Intell. 2026;48(3):2566–80. doi:10.1109/TPAMI.2025.3628027.
8. Rahimi E, Hashemi Nazari SS. A detailed explanation and graphical representation of the Blinder-Oaxaca decomposition method with its application in health inequalities. Emerg Themes Epidemiol. 2021;18(1):12. doi:10.1186/s12982-021-00100-9.
9. Hossain M, Alam MJ, Begum IA, Sun H, McKenzie AM. Blinder-Oaxaca decomposition analysis of the urban-rural gap in child nutrition in Bangladesh. Sci Rep. 2025;15(1):42752. doi:10.1038/s41598-025-26857-6.
10. Penberthy LT, Rivera DR, Lund J, Bruno MA, Meyer AM. An overview of real-world data sources for oncology and considerations for research. CA Cancer J Clin. 2022;72(3):287–300. doi:10.3322/caac.21714.
11. Casella G, Berger R. Statistical inference. 2nd ed. Boca Raton, FL, USA: CRC Press; 2024. doi:10.1201/9781003456285.
12. Kang MJ, Jung KW, Bang SH, Choi SH, Park EH, Yun EH, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2020. Cancer Res Treat. 2023;55(2):385–99. doi:10.4143/crt.2023.447.
13. Park EH, Jung KW, Park NJ, Kang MJ, Yun EH, Kim HJ, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2021. Cancer Res Treat. 2024;56(2):357–71. doi:10.4143/crt.2024.253.
14. Park EH, Jung KW, Park NJ, Kang MJ, Yun EH, Kim HJ, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2022. Cancer Res Treat. 2025;57(2):312–30. doi:10.4143/crt.2025.264.
15. World Bank. Population ages 65 and above (% of total population)—Korea, Rep. New York, NY, USA: UN Population Division; 2024 [cited 2026 Apr 10]. Available from: https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS?locations=KR.
16. Statista. South Korea: cancer crude incidence rate by age, 2022. [cited 2026 Apr 10]. Available from: https://www.statista.com/statistics/1440818/south-korea-cancer-crude-incidence-rate-by-age/.
17. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998;17(19):2265–81. doi:10.1002/(sici)1097-0258(19981015)17:19<2265::aid-sim918>3.0.co;2-b.
18. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. JAMA. 2022;328(24):2446–7. doi:10.1001/jama.2022.21383.
19. Roccetti M. 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):121–34. doi:10.3934/publichealth.2026008.
20. Wang SV, Pottegård A, Crown W, Arlett P, Ashcroft DM, Benchimol EI, et al. Harmonized protocol template to enhance reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: a good practices report of a joint ISPE/ISPOR task force. Pharmacoepidemiol Drug Saf. 2023;32(1):44–55. doi:10.1002/pds.5507.