Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When scalar endpoints cannot identify multidomain stabilization: estimand-aligned trajectory summaries for chronic longitudinal trials

Robert Lundberg

robert.lundberg@liu.se

Chromalytica AB

Thomas Lundeberg

Karolinska Institutet and Sabbatsbergs sjukhus

Johan Dahlén

Linköping University

Research Article

Keywords: Randomized controlled trials, Estimand framework, Longitudinal analysis, Multidomain outcomes, Chronic disease, Clinically important deterioration, Trajectory analysis

Posted Date: August 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10257645/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. During the period in which this work was carried out, Robert Lundberg was employed part-time by Chromalytica AB (developing GC-UV instrumentation in collaboration with the GC-UV Center in Linköping). He holds no financial interest or stock in the company, and the present study has no bearing on the activities of Chromalytica AB. Thomas Lundeberg facilitated Robert Lundberg's work at Linköping University by lending GC-UV and GC-MS instruments; Johan Dahlén supervised the work. None of the authors declare any financial interest in the work or its results.

Abstract

Many chronic conditions involve interacting symptom, functional, and physiological domains measured repeatedly over time. Randomized trials frequently summarize efficacy using baseline-to-endpoint contrasts in a designated primary domain. When the clinical question concerns sustained multidomain stabilization rather than single-domain magnitude change, it is uncertain whether scalar endpoint summaries uniquely determine the estimand of interest. We represent trial outcomes as multidomain longitudinal trajectories and define prespecifiable trajectory-based estimands, including coherent multidomain response, sustained stabilization, and a fragmentation index describing oscillatory temporal organization. We establish structurally that scalar baseline-to-endpoint contrasts are generally non-injective mappings when at least two domains and three measurement occasions are present. We then illustrate the implications using published summary statistics from three chronic randomized trials (anxiety, chronic low back pain, and COPD), and conduct a Monte Carlo study to estimate how often endpoint equivalence can coexist with divergence in trajectory-based estimands under plausible longitudinal dependence structures. All illustrations use published summary statistics only; no individual participant data were accessed. Distinct multidomain longitudinal organizations can produce similar scalar endpoint contrasts while differing in sustained stabilization and risk of short-term deterioration. Across three disease areas, endpoint similarity in a primary domain coexisted with divergence in multidomain response patterns. In simulations reflecting realistic cross-domain correlation and temporal autocorrelation, endpoint equivalence frequently co-occurred with material differences in sustained stabilization or fragmentation. Scalar endpoint analyses remain valid for endpoint-specific estimands. However, when the clinical objective involves sustained multidomain stabilization, endpoint-only contrasts may not uniquely determine longitudinal organization. Trajectory-based estimands can be prespecified within the ICH E9(R1) framework and computed from standard repeated-measure data without altering randomization or intention-to-treat principles. Augmenting endpoint analyses with multidomain longitudinal summaries may improve alignment between the estimand of interest and chronic regulatory disease dynamics.

1. Background

Randomized trials in chronic diseases commonly measure multiple interacting domains over time, including symptoms, functional capacity, physiological markers, and patient-reported health status [1–3]. In chronic obstructive pulmonary disease (COPD), lung function, dyspnoea, and health status evolve jointly and are frequently summarized using composite deterioration endpoints such as clinically important deterioration (CID) [4, 5]. In chronic pain, consensus initiatives such as IMMPACT recommend multidimensional outcome assessment encompassing pain intensity, physical function, emotional functioning, and global improvement [1, 3]. Similar multidomain structures exist in anxiety disorders, where clinician-rated severity, panic symptoms, and functional impairment may change at different rates. Despite this multidimensional measurement, primary efficacy reporting in many trials relies on baseline-to-endpoint contrasts in a single designated domain. This approach is appropriate when the estimand of interest concerns magnitude change in that domain. Under the ICH E9(R1) framework, an estimand

precisely defines the population, variable, intercurrent event strategy, and summary measure of interest [6–8]. More broadly, modern causal inference emphasizes that interpretation of treatment effects depends on explicit specification of the causal quantity being estimated [9]. When the clinical objective involves sustained multidomain stabilization rather than isolated endpoint change, it is unclear whether scalar endpoint contrasts uniquely determine that estimand. From a statistical perspective, reducing a multidomain longitudinal trajectory to a scalar endpoint contrast constitutes a dimensional reduction. While dimensional reduction is often pragmatic, it may not preserve all aspects of temporal and cross-domain organization [10, 11]. If distinct multidomain trajectories can yield similar endpoint contrasts, then endpoint-only evaluation may not uniquely determine longitudinal stabilization patterns.

In this study, we formally examine this structural relationship and assess its practical relevance across three chronic randomized trials. Our objective is not to replace scalar endpoints, but to clarify when augmentation with prespecified multidomain trajectory summaries may improve alignment between estimand definition and chronic disease dynamics.

2. Related work

The present contribution sits at the intersection of three established methodological literatures, and is intended to connect rather than displace them.

2.1 The estimand framework

The ICH E9(R1) addendum formalized the estimand as a precise specification of the population, variable (endpoint), intercurrent event strategy, and population-level summary measure [6]. Subsequent methodological work has broadened this perspective and clarified the relationships among estimands, estimators, and sensitivity analyses [7, 8]. This literature has concentrated largely on how intercurrent events are handled for a given endpoint, rather than on whether a scalar endpoint summary uniquely identifies a multidomain longitudinal objective. We address the latter question directly, treating the choice between scalar and trajectory-based summary measures as itself an estimand specification decision. This perspective aligns with the broader causal-inference literature, which emphasizes that different summaries of the same observed data may correspond to different causal questions and therefore different estimands [9].

2.2 Longitudinal and trajectory modelling

Methods for longitudinal data, mixed-effects models, generalized estimating equations, and latent-class or latent-process trajectory models, are well established [10–13]. These approaches can in principle recover trajectory structure when repeated measurements are modelled jointly. The structural point we make is complementary: when a trial is reported, interpreted, or prespecified through a single baseline-to-endpoint contrast, the trajectory information that longitudinal models are designed to exploit is not identifiable from that scalar summary. The limitation we describe is therefore a property of the summary measure, not of the available modelling machinery.

2.3 Composite and multidomain deterioration endpoints

Composite endpoints are widely used to capture multidomain change, and their interpretation has been critically examined [14, 15]. In COPD, clinically important deterioration (CID) integrates lung function, symptom burden, and exacerbation events into a single multidomain construct [4, 5, 16]. CID demonstrates that the field already recognizes the value of multidomain longitudinal summaries. Our framework generalizes this recognition: we show formally why single-domain endpoint contrasts cannot, in general, recover cross-domain and temporal stabilization, and we cast trajectory-based summaries as estimands that sit within, rather than outside, the established composite-endpoint and ICH E9(R1) logic [6–8].

3. Methods

3.1 Study design and objectives

This is a methodological study examining alignment between scalar endpoint contrasts and multidomain longitudinal stabilization in chronic randomized trials. The study consists of three components: a structural analysis of scalar endpoint reduction under multidomain longitudinal measurement, empirical illustrations using published summary statistics from three chronic randomized trials, and a Monte Carlo simulation study quantifying the frequency of divergence between endpoint contrasts and trajectory-based estimands under plausible chronic-disease dependence structures. No individual participant data (IPD) were accessed. All empirical illustrations were reconstructed using published summary statistics only.

3.2 Conceptual framework: multidomain longitudinal outcomes

We conceptualize each participant's data as a multidomain trajectory observed at baseline and repeated follow-up visits, $Y_{ij(t}$m), for domains j = 1, …, k and occasions m = 0, …, T. Chronic disease trials commonly measure interacting domains (e.g., symptom severity, functional capacity, health status, physiology), as recommended in multidimensional outcome frameworks such as IMMPACT for chronic pain [1, 3] and composite deterioration constructs in COPD (CID) [4, 5, 16].

The standard scalar endpoint contrast in a primary domain is defined as $\Delta EP = Y1(tT) - Y1(t_0)$, where domain 1 is the designated primary variable. We now make explicit the identifying assumption on which the inference from scalar endpoint equivalence to clinical equivalence silently depends, and show it is generally violated.

3.2.1 Non-identifiability of scalar endpoints

Reduction of a multidomain longitudinal trajectory to a scalar endpoint is a projection from a higher-dimensional space to a lower-dimensional summary. Let domains d = 1, …, k (k ≥ 2) be measured at times

$t_0 < t_1 < \ldots < t_T$ ($T \geq 2$), with each domain taking values in a finite totally ordered set (no interval scale assumed), and let a trajectory be the ordered sequence of multidomain configurations. A conventional scalar endpoint depends only on baseline and final values in a designated primary domain, and is therefore invariant to all non-primary domains at every occasion and to the primary domain at intermediate occasions.

Theorem 1

(Non-injectivity of scalar projection). Assume $k \geq 2$ and $T \geq 2$. The mapping from trajectories to scalar endpoints is not injective: there exist distinct trajectories that produce identical scalar endpoint values while differing in sustained coherence or fragmentation. The proof is constructive (Supplement S2). Equivalently, the endpoint map has a nontrivial kernel: by the rank–nullity relation it discards exactly the trajectory directions spanned by non-primary domains and intermediate occasions, so any trajectory functional supported on that kernel is unrecoverable from the endpoint (Supplement S3).

Corollary 1

(Kernel generality). The non-injectivity is a property of the projection, not of any particular summary. Because the projection fibre contains distinct trajectories, any functional non-constant on that fibre is non-identified from the scalar endpoint; including total variation, threshold-crossing counts, time-to-sustained-response, and any other summary of multidomain temporal organization. The result therefore cannot be evaded by substituting a different organizational index; non-identification follows from the dimension reduction itself.

Corollary 2

(Population-level non-identification). Because the projection is many-to-one at the individual level, scalar endpoint distributions under randomized arms do not, in general, identify between-arm contrasts in sustained coherence or fragmentation. Scalar endpoint equivalence therefore does not imply equivalence of multidomain stabilization. For brevity we refer to this situation, defined here as the case in which scalar endpoint comparisons cannot distinguish arms while at least one trajectory-organization estimand differs, as trajectory non-identifiability (TNI). The term is introduced only as a convenient label for this configuration; it denotes a consequence of the projection non-injectivity established above rather than a distinct mathematical object. The claim is existential: it establishes that TNI can occur, not that it occurs with any particular frequency; a separate, quantitative question examined in the Monte Carlo study (§ 3.6, § 4.4).

3.3 Prespecified trajectory-based estimands

To align with the ICH E9(R1) framework [6–8], we define trajectory-based estimands corresponding to clinically meaningful stabilization constructs.

(A) Coherent multidomain response (Δ_C). *Population*: all randomized participants (treatment-policy strategy for intercurrent events). *Variable*: simultaneous attainment of prespecified clinically meaningful improvement in ≥ d domains at the final visit. *Summary*: between-arm difference in probability of coherent response. Clinically meaningful thresholds were anchored to established minimal clinically important differences (MCIDs): TDI ≥ 1 unit [17], applied to the SAC-TDI as its validated self-administered equivalent; CAT ≥ 2 units [18]; pain NRS ≥ 30% or ≥ 2 points [19]; RMDQ minimal clinically important difference [20].

(B) Sustained stabilization (Δ_SC). *Population*: all randomized participants. *Variable*: maintenance of clinically meaningful improvement across consecutive follow-up visits. *Summary*: between-arm difference in probability of sustained stabilization from t*a to t*T.

(C) Fragmentation index (Δ_FI). *Population*: all randomized participants. *Variable*: within-person oscillatory pattern across visits, defined by sign changes in domain-aligned increments. *Summary*: between-arm difference in average fragmentation index. The fragmentation measure is descriptive and intended to capture instability patterns; its specific form is not load-bearing for the identification result.

3.3.1 Formal estimand definitions (ICH E9[R1])

The trajectory-based estimands were defined explicitly in accordance with ICH E9(R1) principles [6–8]. The prespecified estimands and their corresponding populations, variables, intercurrent-event (ICE) strategies, and summary measures are summarized in Table 1.

Table 1
Prespecified estimands defined within the ICH E9(R1) framework.

Estimand	Population	Variable	ICE strategy	Summary measure
Δ_EP (Endpoint contrast)	All randomized participants	Change from baseline to final visit in designated primary domain	Treatment policy	Between-arm mean difference
Δ_C (Coherent multidomain response)	All randomized participants	Simultaneous attainment of prespecified MCID thresholds in ≥ d domains at final visit	Treatment policy	Between-arm difference in probability
Δ_SC (Sustained stabilization)	All randomized participants	Maintenance of MCID-defined improvement across consecutive visits	Treatment policy	Between-arm difference in probability
Δ_FI (Fragmentation index)	All randomized participants	Within-person oscillatory pattern across visits	Treatment policy	Between-arm difference in mean index
Δ_CID (where applicable)	All randomized participants	Time to first clinically important deterioration event	Composite (per trial definition)	Hazard ratio

3.4 Intercurrent events (ICE) strategy

Estimands were defined under a treatment-policy strategy for intercurrent events unless otherwise stated. Intercurrent events (e.g., rescue medication use, discontinuation, exacerbation) were handled as follows: for coherent response and sustained stabilization, intercurrent events were incorporated into outcome trajectories (treatment-policy approach); for CID-based comparisons in COPD, a composite strategy aligned with trial definitions was used [4, 5, 16]; sensitivity analyses exploring hypothetical strategies are described in Supplement S4. This mapping follows ICH E9(R1) principles for transparent estimand definition [6–8]. Table 2 summarizes the intercurrent-event strategies applied to each estimand under the proposed framework.

Table 2
Mapping of intercurrent-event strategies across prespecified estimands.

Intercurrent event	Δ_EP	Δ_C	Δ_SC	Δ_FI	Δ_CID
Treatment discontinuation	Treatment policy	Treatment policy	Treatment policy	Treatment policy	Treatment policy
Rescue medication use	Included in trajectory	Included in trajectory	Included in trajectory	Included in trajectory	Included per composite definition
Moderate exacerbation	Included	Included	Included	Included	Event component
Severe exacerbation	Included	Included	Included	Included	Event component
Missing outcome data	Mixed model under MAR	Multiple-imputation sensitivity	Multiple-imputation sensitivity	Complete-case + sensitivity	Standard survival censoring

3.5 Empirical illustrations

Three published chronic randomized trials were used as empirical illustrations: a trial of mindfulness versus escitalopram for anxiety [21]; a trial of virtual yoga for chronic low back pain [22]; and the EMAX trial of dual bronchodilation versus monotherapy in COPD [2]. For each trial, arm-level summary statistics (means, standard deviations, and sample sizes, or between-group mean differences) were extracted directly from the published reports. These published values are used as illustrations of the structural argument; we do not reconstruct individual participant data or attempt to recover original joint distributions, which are not identifiable from marginal summaries.

3.6 Monte Carlo study

A Monte Carlo study estimated how frequently scalar endpoint equivalence coexists with divergence in trajectory-based parameters under plausible chronic-trial dependence structures. The simulation grid varied domain dimensionality ($k \in (2, 3, 4)$), number of visits ($T + 1 \in (3, 4, 6)$), temporal autocorrelation (ρ), and cross-domain equicorrelation (α), with multidomain latent deviations generated under an AR(1) temporal structure. Endpoint equivalence was defined by $|\Delta EP| \le \delta$ and divergence by $|\Delta \mathrm{EP}| \le \delta$ together with $|\Delta SC| \ge \tau \mathrm{SC}$ or $|\Delta FI| \ge \tau \mathrm{FI}$. Full specification, including thresholds, replicate counts, and confidence-interval construction, is given in Supplement S4. Simulation design followed established guidance for statistical simulation studies [23, 24].

4. Results

4.1 Scalar endpoint contrasts do not uniquely determine multidomain stabilization

Across all three trials, scalar baseline-to-endpoint contrasts summarized treatment effects within a designated primary domain. In the anxiety trial, the two arms occupied the same primary-domain severity at the week-8 endpoint (CGI-S 3.09 vs 3.09; between-arm difference 0.00; SE 0.15; P = .98) and were numerically identical again at week 24 (2.92 vs 2.92); the change-from-baseline contrast met the prespecified noninferiority margin of − 0.495 (Δ_EP = − 0.07; SE 0.16; 95% CI − 0.38 to 0.23; P = .65) [21]. In chronic low back pain, pain reduction at weeks 12 and 24 showed clear between-group differences [22]. In COPD (EMAX), trough FEV_1 improved significantly with dual bronchodilation compared with monotherapy [2]. Viewed independently, these scalar contrasts appear to summarize efficacy adequately within their respective primary domains. However, scalar endpoint contrasts do not encode cross-domain or temporal organization (Fig. 1).

4.2 Multidomain response patterns diverge beyond scalar endpoints

When additional domains are considered jointly, divergence emerges. In the anxiety trial, clinician- and patient-rated secondary domains showed heterogeneous response patterns despite similar CGI-S change [21]. In chronic low back pain, reductions in analgesic use and improvements in sleep quality evolved alongside pain reduction [22]. In COPD, dual therapy reduced the probability of clinically important deterioration (CID) across all three prespecified composite definitions, each integrating lung function with symptom, health-status and exacerbation components rather than lung function alone [2, 4, 5, 16]. These patterns indicate that endpoint contrasts in a single domain do not uniquely determine coherent multidomain response probabilities (Fig. 2).

4.3 Temporal stabilization reveals additional structure

Longitudinal organization across visits further differentiates treatment effects. In the anxiety trial, although the arms reached equivalent primary-domain severity at weeks 8 and 24, they did so via different intermediate trajectories: escitalopram improved faster early, both in completer group means (week-4 CGI-S 3.34 vs 3.42 for MBSR) and in the adjusted intent-to-treat mixed model (week-4 between-arm difference − 0.20; 95% CI − 0.43 to 0.04), before the arms converged to indistinguishable endpoints. This early divergence was more pronounced on the patient-rated OASIS, where escitalopram showed a statistically significant week-4 advantage (adjusted difference − 1.2; 95% CI − 2.02 to − 0.35; P = .01) that had closed by the week-8 endpoint (− 0.7; 95% CI − 1.51 to 0.17; P = .12) and remained non-significant at week 24 (− 0.5; P = .28). In chronic low back pain, improvements observed at week 12 were sustained at week 24 across pain, function, and sleep domains [22]. In COPD, CID time-to-event curves demonstrated reduced short-term deterioration risk under dual therapy [2]. Similar observations from adherence studies suggest that longitudinal maintenance of disease control is associated with clinically important outcomes such as hospitalization and mortality, emphasizing the importance of sustained stabilization rather than isolated endpoint assessment [25]. Temporal stabilization patterns differ even when scalar endpoint magnitude appears modest (Fig. 3).

4.4 Frequency of divergence under realistic trial structures

Monte Carlo simulations showed that scalar endpoint equivalence coexisted with divergence in trajectory-based estimands across the full grid, but that divergence frequency depended strongly on the number of measurement occasions. Divergence was most frequent in shorter longitudinal structures with fewer visits (three visits: $p_D \approx 0.52–0.57$ across domain counts) and decreased sharply as the number of visits increased (four visits: $p_D \approx 0.08–0.11$; six visits: $p_D \approx 0.03–0.05$). This decline is expected under the sustained-stabilization definition, which requires maintenance of clinically meaningful improvement across consecutive follow-up visits: as the number of visits grows, sustained stabilization becomes a stricter condition, and a transient departure from the responder state is more readily absorbed by a longer sequence of maintained visits, so fewer trajectories are classified as divergent. Domain count had little systematic effect at any schedule. Endpoint–trajectory discordance is therefore most pronounced in short chronic trials with limited repeated measurement, whereas richer follow-up schedules reduce hidden divergence by more tightly constraining stabilization patterns. The qualitative pattern, divergence being the highest with few visits and declining as visits increase, was invariant across all threshold and dependence settings examined; its magnitude depended on the equivalence margin δ, with full sensitivity analyses reported in Supplement S4 (Fig. 4).

4.5 Frequency of endpoint–trajectory divergence

The Monte Carlo study quantified the frequency with which scalar endpoint equivalence coexisted with divergence in trajectory-based estimands under realistic chronic-trial conditions. Under the primary specification (moderate cross-domain and temporal dependence, $\rho = 0.6$, $\alpha = 0.5$), divergence ranged from roughly 0.03 in the richest six-visit schedules to about 0.57 in the shortest three-visit trials, and was consistently substantial in the short-trial regime. Sensitivity analyses varying the divergence thresholds and dependence parameters are reported in Supplement S4; the direction of the visit-count effect was invariant across all settings examined, while its magnitude scaled with the equivalence margin. These results indicate that the structural non-identification demonstrated analytically is not a rare mathematical artifact, but emerges under plausible chronic longitudinal trial conditions, and is most pronounced precisely where repeated measurement is sparsest.

4.6 Cross-disease consistency

Across the three empirical illustrations, anxiety, chronic low back pain, and COPD, the same structural pattern was observed: scalar endpoint contrasts summarized treatment effects within a primary domain; additional domains revealed distinct cross-domain response configurations; repeated measurement enabled differentiation of sustained stabilization versus transient improvement; and composite or trajectory-based estimands captured longitudinal organization not uniquely determined by scalar endpoints. The COPD EMAX trial is particularly instructive, as CID analyses, constructed prospectively as multidomain deterioration endpoints, demonstrated divergence even when selected scalar health-status contrasts were modest [2]. The consistency of this pattern across distinct chronic disease systems

supports the general structural interpretation: in multidomain longitudinal settings, endpoint-only reduction does not uniquely determine multidomain stabilization.

5. Discussion

5.1 What this study clarifies and what it does not

This study does not question the validity of scalar endpoint analyses. Scalar endpoints remain appropriate when the estimand of interest concerns change in a single domain at a specified timepoint. Rather, we clarify a structural property of multidomain longitudinal trials: when outcomes are measured across multiple interacting domains and over repeated visits, reduction to a scalar baseline-to-endpoint contrast does not, in general, uniquely determine longitudinal stabilization patterns.

The distinction is therefore not between “valid” and “invalid” endpoints, but between different estimands. Endpoint contrasts validly estimate endpoint-specific quantities. However, when the clinical objective concerns sustained multidomain stabilization, coordinated improvement across symptom, functional, and physiological domains, the endpoint-only summary may not uniquely determine that estimand. This limitation arises from dimensional reduction of multidomain longitudinal data structures [10, 11], not from inadequate power, inappropriate modeling, or scale properties.

5.2 Structural clarification in the context of estimands

Under the ICH E9(R1) framework, precise estimand definition requires explicit specification of the population, variable, intercurrent event strategy, and summary measure [6–8]. When the variable of interest is multidomain longitudinal stabilization rather than single-domain magnitude change, the summary measure must preserve cross-domain and temporal structure. This requirement is consistent with causal-inference principles that distinguish the scientific question (estimand) from the statistical procedure used to estimate it [9].

Our structural analysis shows that when at least two clinically relevant domains and at least three measurement occasions are present, scalar baseline-to-endpoint contrasts constitute a non-injective mapping from trajectory space to endpoint space. Consequently, distinct multidomain trajectories may produce identical scalar endpoint contrasts. This structural property holds irrespective of whether outcomes are treated as ordinal or numerical (Supplement S1–S3). From a longitudinal modeling perspective, repeated-measures data encode trajectory organization that is not recoverable from baseline and endpoint alone [10, 11, 13]. The implication is not that endpoint analyses are incorrect, but that they answer a narrower question than trajectory-based estimands designed to capture sustained stabilization.

5.3 Empirical illustrations across chronic regulatory systems

The structural property was observed consistently across three chronic regulatory systems: anxiety disorders, chronic low back pain, and COPD. In anxiety, clinician-rated severity (CGI-S) change did not significantly differ between arms, yet panic symptom improvement (PDSS) and other domains exhibited heterogeneous response patterns [21]. In chronic low back pain, pain reduction alone did not uniquely characterize the joint evolution of functional disability, sleep quality, and medication use [22]. In COPD, dual bronchodilation reduced the probability of clinically important deterioration (CID) across multiple composite definitions, including definitions not solely driven by lung function [2, 4, 5, 16].

The COPD literature already recognizes that multidomain deterioration constructs such as CID may better reflect disease instability than single-domain contrasts [4, 5, 16]. This perspective is consistent with evidence that adherence-related differences in disease control translate into downstream differences in hospitalization and mortality risk [25]. Our framework generalizes this recognition by formally demonstrating that scalar endpoint contrasts do not uniquely determine cross-domain and temporal stabilization, even when composite deterioration endpoints are not explicitly analyzed. The consistency of this pattern across psychiatric, musculoskeletal, and pulmonary disease contexts suggests that the identification issue is structural rather than disease-specific. Chronic diseases are multidomain regulatory systems in which symptom burden, function, and physiological measures interact dynamically over time [26–28].

5.4 Frequency under plausible longitudinal structures

The Monte Carlo analysis demonstrates that divergence between scalar endpoint equivalence and trajectory-based estimands is not a rare theoretical artifact. Under plausible chronic-trial conditions, moderate cross-domain correlation, moderate temporal autocorrelation, and three or more measurement occasions, endpoint equivalence frequently coexisted with material differences in sustained stabilization or fragmentation estimands. Simulation design followed established methodological guidance for statistical simulation studies [23, 24]. Divergence probability was highest in short longitudinal structures with few visits and declined as the number of measurement occasions increased, because sustained stabilization becomes a stricter condition when maintenance is required across more consecutive visits; domain count had comparatively little systematic effect. Trials measuring interacting domains but with sparse repeated measurement are therefore the setting in which endpoint-only summaries are most likely to obscure trajectory divergence, whereas richer follow-up schedules more tightly constrain stabilization patterns.

5.5 Relationship to composite endpoints

Composite endpoints are widely used in chronic disease trials to capture multidomain events [14, 15]. However, methodological concerns regarding interpretation of composite endpoints have also been well documented [14, 15]. The present framework does not propose replacing composite endpoints but clarifies the structural relationship between scalar reduction and multidomain stabilization. CID in COPD, for example, integrates lung function, symptom burden, and exacerbation events [4, 5, 16]. The fact that CID divergence in EMAX held across all three composite definitions, none of which is driven by FEV_1

alone, illustrates that multidomain stabilization cannot be inferred solely from a single-domain endpoint contrast [2].

The multidomain information discarded by scalar reduction is not merely additional descriptive detail: in its most developed instance it carries external prognostic weight. Longitudinal maintenance of disease control has been associated with downstream differences in hospitalization and mortality [25], and clinically important deterioration, a sustained-deterioration construct integrating lung function, symptom burden, and exacerbation events, is the composite through which that association is operationalized [4, 5, 16, 25]. Sustained stabilization (Δ_{SC}) generalizes the same logic, summarizing whether multidomain improvement is maintained across consecutive visits rather than attained momentarily at a single endpoint. We do not claim to demonstrate that Δ_{SC} predicts hospitalization or mortality in the present illustrations; that would require individual participant data that were not accessed. We claim the narrower, verifiable point: the class of sustained-stabilization summaries to which Δ_{SC} belongs contains constructs that are externally validated against hard outcomes, so the structural non-identification established here represents loss of longitudinal signal with independent prognostic standing rather than information loss in the abstract.

Thus, trajectory-based estimands should be viewed as aligned extensions within established composite-endpoint logic, framed explicitly under the estimand structure of ICH E9(R1) [6–8].

5.6 Boundary conditions

It is essential to define when endpoint-only evaluation is sufficient. Endpoint-only summaries are adequate when the outcome space is effectively one-dimensional; when intermediate temporal organization is not clinically meaningful; when the estimand explicitly concerns magnitude change at a prespecified timepoint; and when cross-domain coherence is not part of the clinical objective. In such settings, augmentation with trajectory-based summaries is unnecessary. The present framework applies specifically to chronic multidomain longitudinal systems in which sustained stabilization across domains is clinically relevant.

5.7 Implementation within standard trial practice

Trajectory-based estimands can be incorporated without altering randomization, allocation concealment, or intention-to-treat principles. Operational requirements are modest: repeated measurement across at least two follow-up visits; prespecification of clinically meaningful thresholds anchored in established MCIDs [17–20]; transparent intercurrent-event handling consistent with ICH E9(R1) [6–8]; and standard regression or survival modeling approaches [10, 11, 29]. Mixed-effects models and Cox proportional-hazards models are sufficient to estimate the proposed estimands [10, 11, 29]. No specialized modeling framework is required. Trajectory-based summaries can be designated as key secondary or supportive estimands within hierarchical testing structures. Their inclusion strengthens alignment between estimand definition and chronic disease dynamics rather than replacing established primary endpoints.

5.8 Limitations

This study is methodological and relies on reconstruction from published summary statistics. No individual participant data were accessed. Although reconstruction procedures were constrained to reproduce reported marginal means and dispersion, they cannot recover original joint distributions. The Monte Carlo analysis explores plausible longitudinal structures rather than empirically observed distributions across disease areas; divergence frequency estimates are therefore conditional on specified parameter ranges and divergence thresholds. The fragmentation index is a descriptive measure of temporal organization and does not constitute a validated clinical severity scale. Its specific functional form, sign changes in domain-aligned increments, is one of several plausible operationalizations; total variation, the number of turning points, switching frequency, or spectral summaries of the trajectory would serve comparably, and it is offered as a descriptive illustration rather than a validated instability metric. The structural identification result, and the sustained-stabilization findings that carry the primary quantitative claims, are independent of this choice. Finally, the empirical illustrations do not reinterpret original trial conclusions; they demonstrate structural compatibility of endpoint similarity with multidomain divergence, not clinical superiority of any intervention.

5.9 Overall interpretation

Scalar endpoint contrasts validly estimate endpoint-specific estimands. However, when the clinical objective involves sustained multidomain stabilization in chronic longitudinal systems, endpoint-only reduction may not uniquely determine longitudinal organization. The contribution of this study is threefold: structural clarification of non-identifiability under scalar reduction, empirical illustration across distinct chronic disease systems, and quantitative demonstration of divergence frequency under realistic trial structures. The goal is refinement of estimand alignment; not replacement of established endpoint methodology.

6. Conclusions

Scalar baseline-to-endpoint contrasts validly estimate endpoint-specific treatment effects. However, in chronic multidomain longitudinal trials, reduction to a single scalar endpoint does not, in general, uniquely determine cross-domain and temporal stabilization patterns. When the clinical objective concerns sustained multidomain stabilization rather than isolated magnitude change at a specific visit, endpoint-only summaries may not fully align with the estimand of interest. This structural limitation arises from dimensional reduction of multidomain longitudinal trajectories and is independent of scale type or modeling approach.

Across anxiety, chronic low back pain, and COPD trials, empirical illustrations demonstrated that similar scalar endpoint contrasts can coexist with divergent multidomain response configurations and temporal stabilization patterns. Monte Carlo analyses further showed that such divergence is not rare under plausible chronic-trial dependence structures. Trajectory-based estimands, such as coherent multidomain response, sustained stabilization, and explicitly defined composite deterioration measures, can be prespecified within the ICH E9(R1) framework and estimated using standard repeated-measures

data without altering randomization or intention-to-treat principles. The objective is refinement of estimand alignment rather than replacement of established endpoint methodology.

Declarations

Ethics approval and consent to participate

Not applicable. This methodological study used only published, aggregate summary statistics; no individual participant data were accessed.

Consent for publication

Not applicable.

Competing Interests

During the period in which this work was carried out, Robert Lundberg was employed part-time by Chromalytica AB (developing GC-UV instrumentation in collaboration with the GC-UV Center in Linköping). He holds no financial interest or stock in the company, and the present study has no bearing on the activities of Chromalytica AB. Thomas Lundeberg facilitated Robert Lundberg's work at Linköping University by lending GC-UV and GC-MS instruments; Johan Dahlén supervised the work. None of the authors declare any financial interest in the work or its results.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The work was conducted as a research collaboration between Linköping University and Karolinska Institutet / Sabbatsbergs sjukhus, Stockholm.

Author Contribution

Robert Lundberg: Writing – Original Draft, Writing – review & editing. Thomas Lundeberg: Conceptualization, Formal analysis, Investigation, Validation, Visualization, Writing – Original Draft, Writing – review & editing. Johan Dahlén: Project administration, Supervision, Writing – review & editing.

Acknowledgement

This work was conducted as a research collaboration between Linköping University and Karolinska Institutet / Sabbatsbergs sjukhus, Stockholm. Johan Lundeberg, Skånes universitetssjukhus, Lund, is acknowledged for his supportive input.

Availability of data and materials

All summary statistics used as calibration targets are reported in the published trials cited [2, 21, 22] and are available in the cited published reports. Simulation code and the reproducibility manifest for the Monte Carlo analysis (random seeds, parameter grid, and sensitivity outputs) are provided as an additional reproducibility file and will be deposited in a public repository on acceptance.

References

1. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(1–2):9–19. 10.1016/j.pain.2004.09.012.
2. Maltais F, Bjermer L, Kerwin EM, Jones PW, Watkins ML, Tombs L, et al. Efficacy of umeclidinium/vilanterol versus umeclidinium and salmeterol monotherapies in symptomatic patients with COPD not receiving inhaled corticosteroids: the EMAX randomised trial. Respir Res. 2019;20(1):238. 10.1186/s12931-019-1193-9.
3. Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. Pain. 2008;137(2):276–85. 10.1016/j.pain.2007.09.002.
4. Singh D, Maleki-Yazdi MR, Tombs L, Iqbal A, Fahy WA, Naya I. Prevention of clinically important deterioration in COPD with umeclidinium/vilanterol: a post hoc analysis. Int J Chron Obstruct Pulmon Dis. 2016;11:1413–24. 10.2147/COPD.S103212.
5. Naya I, Tombs L, Lipson DA, Compton C. Long-term outcomes following first short-term clinically important deterioration in COPD. Respir Res. 2018;19(1):222. 10.1186/s12931-018-0916-3.
6. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Harmonised Guideline E9(R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials. Geneva: ICH; 2019.
7. Mallinckrodt C, Molenberghs G, Lipkovich I, Ratitch B. Estimands, Estimators and Sensitivity Analysis in Clinical Trials. Boca Raton: CRC; 2019. 10.1201/9780429488825.
8. Akacha M, Bretz F, Ruberg S. Estimands in clinical trials – broadening the perspective. Stat Med. 2017;36(1):5–19. 10.1002/sim.7223.
9. Pearl J. Causal inference in statistics: an overview. Stat Surv. 2009;3:96–146. 10.1214/09-SS057.
10. Diggle PJ, Heagerty P, Liang KY, Zeger SL. Analysis of Longitudinal Data. 2nd ed. Oxford: Oxford University Press; 2002.
11. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. 2nd ed. Hoboken: Wiley; 2011.
12. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford: Oxford University Press; 2003.

13. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. J Stat Softw. 2017;78(2):1–56. 10.18637/jss.v078.i02.
14. Ferreira-González I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, et al. Methodologic discussions for using and interpreting composite endpoints are limited but still identify major concerns. J Clin Epidemiol. 2007;60(7):651–7. 10.1016/j.jclinepi.2006.10.020.
15. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. BMJ. 2010;341:c3920. 10.1136/bmj.c3920.
16. Anzueto AR, Vogelmeier CF, Kostikas K, Mezzi K, Fucile S, Bader G, et al. The effect of indacaterol/glycopyrronium versus tiotropium or salmeterol/fluticasone on the prevention of clinically important deterioration in COPD. Int J Chron Obstruct Pulmon Dis. 2017;12:1325–37. 10.2147/COPD.S130867.
17. Mahler DA, Witek TJ Jr. The MCID of the transition dyspnea index is a total score of one unit. COPD. 2005;2(1):99–103. 10.1081/COPD-200050666.
18. Kon SSC, Canavan JL, Jones SE, Nolan CM, Clark AL, Dickson MJ, et al. Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. Lancet Respir Med. 2014;2(3):195–203. 10.1016/S2213-2600(14)70001-3.
19. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001;94(2):149–58. 10.1016/S0304-3959(01)00349-9.
20. Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. J Clin Epidemiol. 2006;59(1):45–52. 10.1016/j.jclinepi.2005.03.018.
21. Hoge EA, Bui E, Mete M, Dutton MA, Baker AW, Simon NM. Mindfulness-based stress reduction vs escitalopram for the treatment of adults with anxiety disorders: a randomized clinical trial. JAMA Psychiatry. 2023;80(1):13–21. 10.1001/jamapsychiatry.2022.3679.
22. Tankha H, Gaskins D, Shallcross A, Rothberg M, Hu B, Guo N, et al. Effectiveness of virtual yoga for chronic low back pain: a randomized clinical trial. JAMA Netw Open. 2024;7(11):e2442339. 10.1001/jamanetworkopen.2024.42339.
23. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38(11):2074–102. 10.1002/sim.8086.
24. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Stat Med. 2006;25(24):4279–92. 10.1002/sim.2673.
25. Vestbo J, Anderson JA, Calverley PMA, Celli B, Ferguson GT, Jenkins C, et al. Adherence to inhaled therapy, mortality and hospital admission in COPD. Thorax. 2009;64(11):939–43. 10.1136/thx.2009.113662.
26. Agusti A, Bel E, Thomas M, Vogelmeier C, Brusselle G, Holgate S, et al. Treatable traits: toward precision medicine of chronic airway diseases. Eur Respir J. 2016;47(2):410–19.

10.1183/13993003.01359-2015.

27. Celli BR, Wedzicha JA. Update on clinical aspects of chronic obstructive pulmonary disease. N Engl J Med. 2019;381(13):1257–66. 10.1056/NEJMra1900500.
28. Turk DC, Wilson HD, Cahana A. Treatment of chronic non-cancer pain. Lancet. 2011;377(9784):2226–35. 10.1016/S0140-6736(11)60402-9.
29. Cox DR. Regression models and life-tables. J R Stat Soc Ser B. 1972;34(2):187–220. 10.1111/j.2517-6161.1972.tb00899.x.

Figures

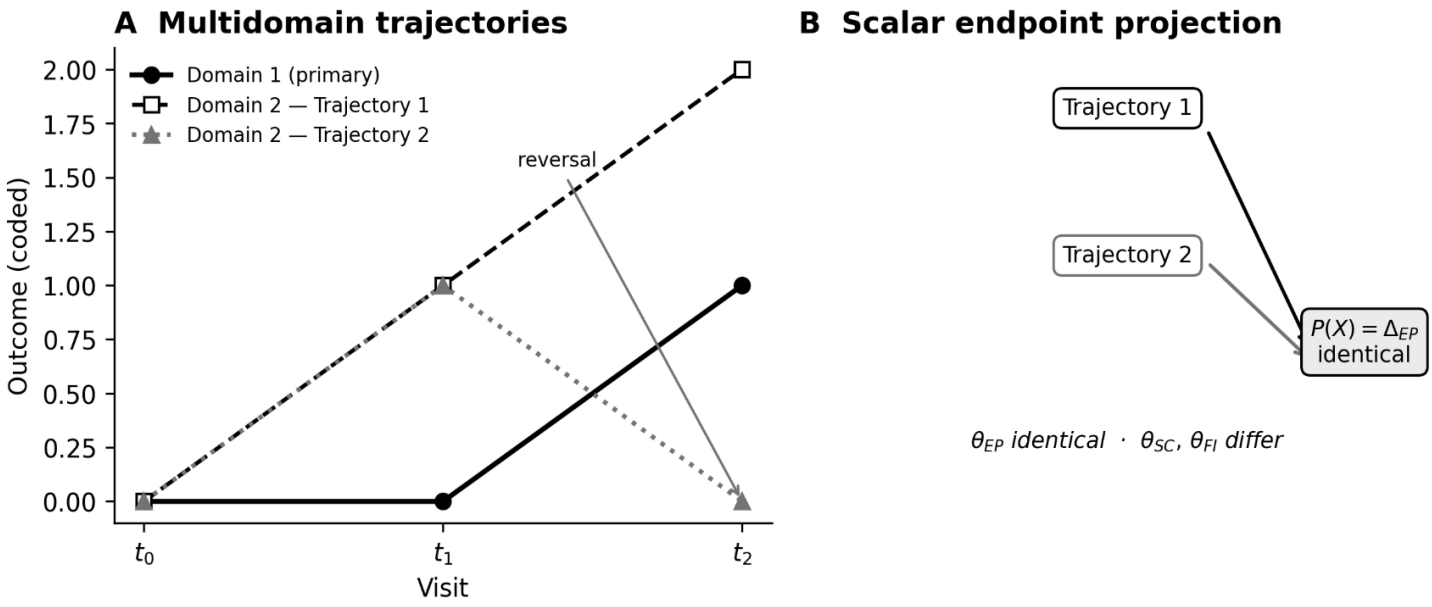


Figure 1

Projection and non-identifiability concept. (A) Two multidomain trajectories share an identical primary-domain (Domain 1) baseline and endpoint but differ in the temporal organization of a secondary domain (Domain 2): one is monotone, the other reverses at the final visit. (B) Both trajectories map to the same scalar endpoint contrast P(X) = Δ_EP, so θ_EP is identical while the sustained-stabilization (θ_SC) and fragmentation (θ_FI) parameters differ.

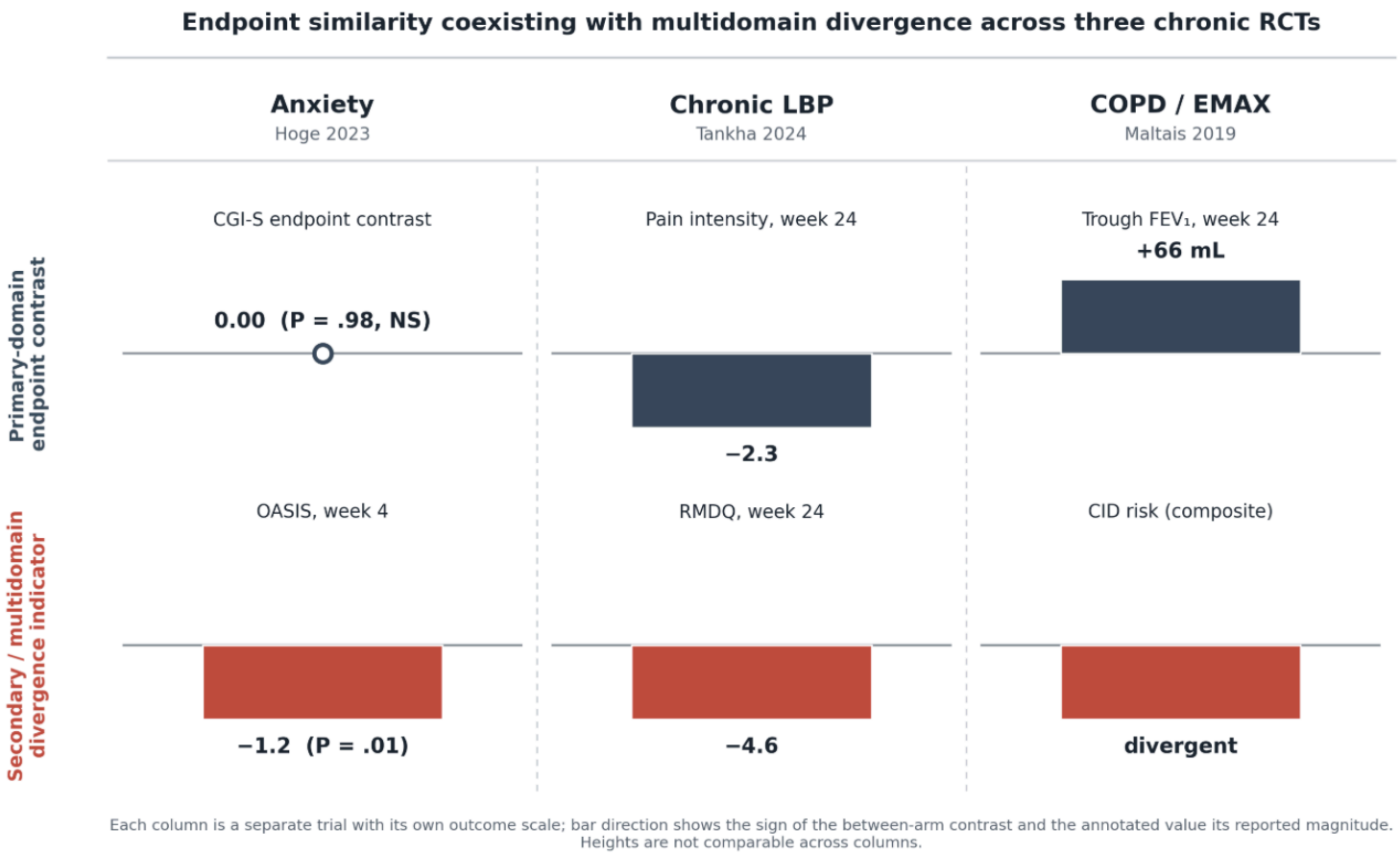


Figure 2

Endpoint similarity coexisting with multidomain divergence across three chronic RCTs. Top row: the primary-domain endpoint contrast for each trial (anxiety, chronic low back pain, COPD/EMAX). Bottom row: a secondary or multidomain divergence indicator from the same trial. Values are reported between-arm contrasts from the published trials: anxiety, week-8 CGI-S endpoint difference (0.00, non-significant) and the significant week-4 OASIS separation (−1.2; P = .01) that closes by endpoint; chronic low back pain, week-24 pain intensity (−2.3) and Roland–Morris disability (−4.6); COPD, week-24 trough FEV_1 (+66 mL, umeclidinium/vilanterol vs umeclidinium) and divergent clinically important deterioration (CID) risk. Each column has its own outcome scale; bar direction indicates the sign of the contrast and the annotated value its reported magnitude, and heights are not comparable across columns.

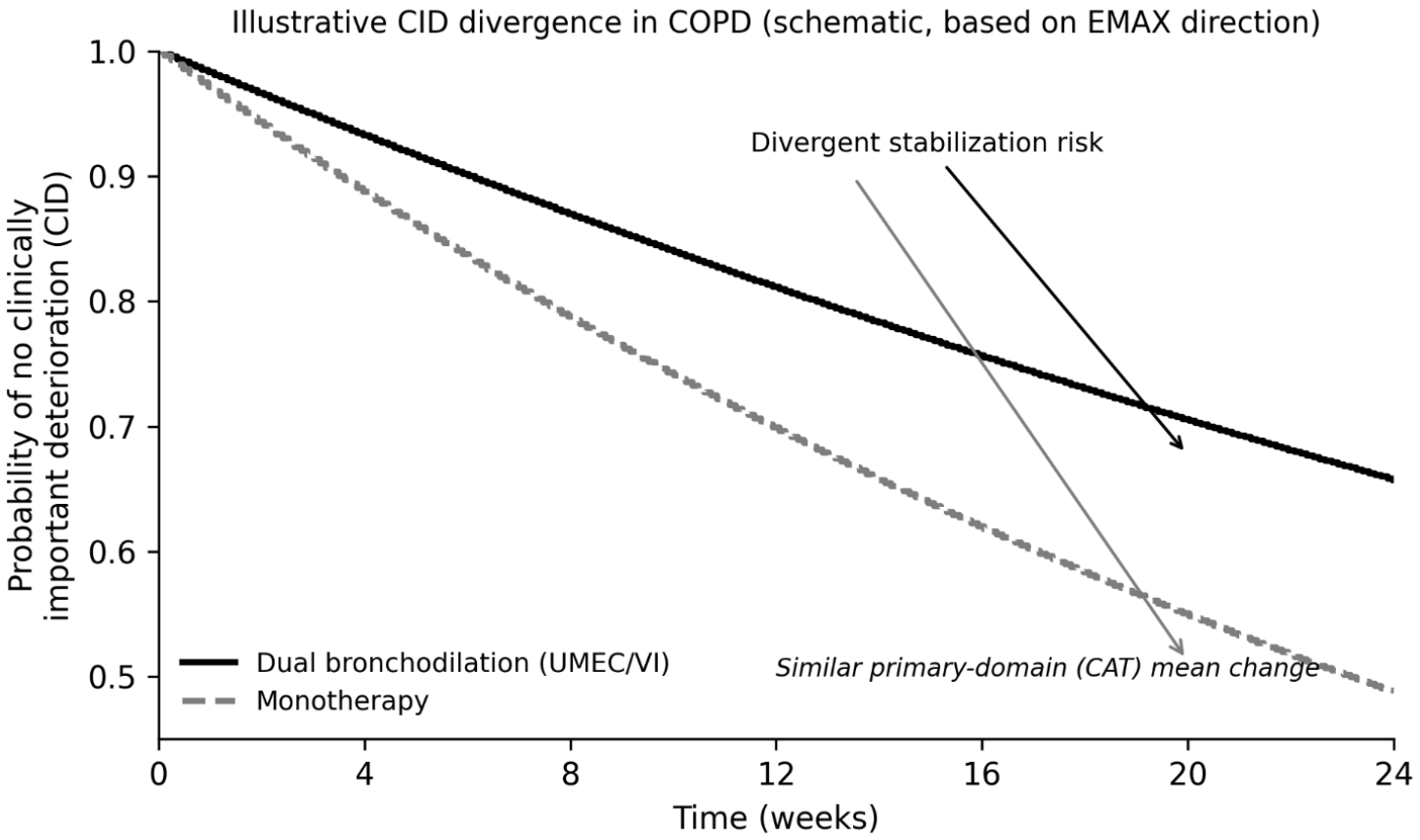


Figure 3

Illustrative CID divergence in COPD. Schematic, simplified time-to-event curves for the probability of remaining free of clinically important deterioration (CID) under dual bronchodilation versus monotherapy, reproducing the direction of effect reported in EMAX [2]. The illustration shows how divergent stabilization risk can coexist with a modest difference in a single primary-domain (CAT) contrast.

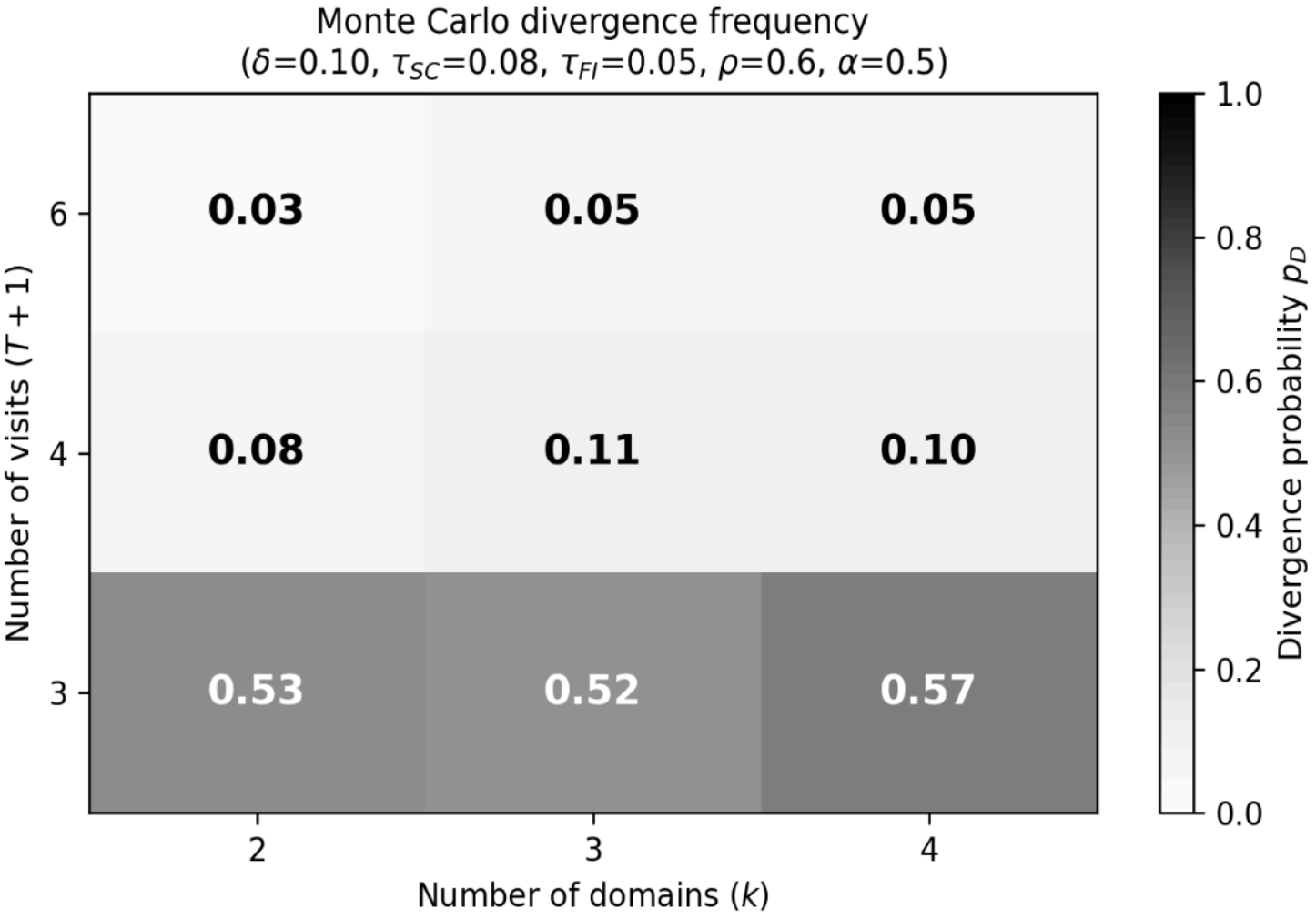


Figure 4

Monte Carlo divergence frequency. Estimated probability p_D that scalar endpoint equivalence (|Δ_EP| ≤ δ) coexists with divergence in a trajectory-based estimand (|Δ_SC| ≥ τ_SC or |Δ_FI| ≥ τ_FI), across a grid of domain count (k) and number of visits (T + 1). Values are Monte Carlo estimates under the primary specification (δ = 0.10, τ_SC = 0.08, τ_FI = 0.05, ρ = 0.6, α = 0.5; 800 replicates per cell, fixed seed). Full specification, seeds, and threshold and dependence-parameter sensitivity analyses are in Supplement S4.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BMCMRMSupplemental.docx