Supplementary Information

The Dark Matter of Clinical Data: an identification limit on disentangling physiology from care process in electronic health records

This Supplementary Information (SI) contains the material that supports, but is not required to follow, the main text: notation and definitions in full (S1), an extended review of related work (S2), complete proofs and additional results (S3 and S4), the full derivation of the worked example with an extended sensitivity analysis (S5), and an extended discussion (S6). Theorem, lemma, corollary, proposition, remark, and assumption numbers match the main text; SI Theorem 1 adds part (iv), SI Section S4 gives remarks on Proposition 4 and proofs and checks for Proposition 5 and Corollary 3, and Corollary 2 and Remarks 1 and 2 appear only here. Equations here are numbered (S1) to (S8); Figures and Tables prefixed "S" are in this document, and all others refer to the main text. References are numbered independently for this document.

S1 Notation, kinds of absence, and definitions in full

We work on a probability space $(\Omega, \mathcal{F}, P)$ supporting the variables in Table S1. Lower-case letters denote values; $p(\cdot)$ denotes a probability mass function or density, as appropriate. For random elements X, W, U, the conditional independence $X \perp\!\!\!\perp W \mid U$ means $p(x \mid w, u) = p(x \mid u)$ for all (w, u) in the support of (W, U); all conditional distributions are assumed to exist as regular conditional distributions. The environment E takes values in a finite set $\mathcal{E}$ with $p(e) > 0$ for every $e \in \mathcal{E}$. We write

$$\mathcal{Z}_{\text{sh}} := \{\, z : p(z \mid E = e) > 0 \text{ for every } e \in \mathcal{E} \,\} \tag{S1}$$

for the *shared support* of the representation across environments; statements about cross-environment invariance are made only on $\mathcal{Z}_{\text{sh}}$.

Table S1: Variables, their interpretation, and their observational status.

Symbol	Interpretation	Status
$Z^\star$	True, complete physiological state of the patient.	Latent; never observed.
Z	Learned representation intended as a proxy for $Z^\star$ (e.g., a transformer embedding), computed by a fixed encoder from the observed record.	Observed (constructed from data).
E	Clinical environment (hospital or health system).	Observed.

Symbol	Interpretation	Status
C	Censoring before measurement: death, discharge, or transition to comfort care.	Observed.
$S^\star$	Clinician's ordering intent for the test.	Latent.
S	Measurement indicator, $S = S^\star(1 - C)$: the test was performed.	Observed; deterministic given $(S^\star, C)$.
V	Recorded laboratory value, defined only on the event $\{S = 1\}$.	Observed when $S = 1$; undefined when $S = 0$.
V°	Analyte level: the value the assay would return, a component of $Z^\star$; in m-graph terms, a partially observed variable with missingness mechanism $R_{V^\circ} = 1 - S$.	Latent; equals V on $\{S = 1\}$.
$\tilde{V}$	Analyst's sentinel encoding of V on the whole population (Eq. (S2)).	Constructed by the analyst.
Y	Outcome of interest (e.g., mortality, sepsis, a later laboratory abnormality).	Observed.

Two conventions deserve emphasis. First, Z is treated as a fixed measurable function of the observed record, as when an encoder is trained on separate data and then frozen; the results concern what Z can reveal, not how it is estimated. Second, the physiological quantity that a laboratory test would assay (for example, the circulating lactate concentration) is part of $Z^\star$ and exists whether or not a test is ordered. The variable V is different: it is the output of a particular measurement event, produced by a specimen, an assay, and a time stamp. When no test is performed, there is no such event and therefore no value of V.

S1.1 Kinds of absence

Much of the confusion about "missing" laboratory data comes from treating four different situations as one. Table S2 separates them, and main Figure 2 illustrates the case that matters most here.

Table S2: Four kinds of absence in electronic health record (EHR) data, and why they differ formally.

Kind	Example	Does the quantity exist?	Formal status	Consequence for analysis
Unobserved (latent)	True circulating lactate concentration; true sepsis state	Yes	Component of $Z^\star$; a random variable on all of Ω that is never recorded	Must be reasoned about through proxies (Z, V)

Kind	Example	Does the quantity exist?	Formal status	Consequence for analysis
Missing	Result lost to an interface error after the assay was run	Yes	V is defined but unrecorded; standard missing-data theory applies [1]	Imputation targets a real, if unseen, value
Structurally undefined	Lactate never ordered, or patient died before the draw	No	V is not a random variable on $\{S = 0\}$ (Assumption 3); a "value had it been measured" is a counterfactual that must be *posited*	Any value placed on $\{S = 0\}$ is a modeling choice, not an estimate
Sentinel-coded	"Not ordered" stored as -1 or as a missingness indicator	Constructed	$\tilde{V}$ is a deterministic function of (S, V) defined everywhere (Eq. (S2))	Encodes S exactly (Theorem 1)

The structurally undefined case is the one created by selective ordering, and it is what makes a sentinel code more than a harmless placeholder. If $\varnothing$ is any symbol outside the range of V (for lactate, -1 serves), the sentinel encoding is

$$\tilde{V} := \begin{cases} V & \text{on } \{S = 1\}, \\ \varnothing & \text{on } \{S = 0\}. \end{cases} \tag{S2}$$

Because $\tilde{V} = \varnothing$ exactly when the test was not performed, the encoded value contains the ordering indicator itself. If a site orders lactate only when a clinician suspects sepsis, a code of -1 records the clinician's lack of suspicion, not anything about lactate (main Figure 2b).

S1.2 Notes on the assumptions

Censoring and mortality outcomes. When Y is mortality, a death before the specimen could be drawn forces both $S = 0$ and $Y = 1$, so the ordering indicator carries outcome information beyond physiology and Assumption 6a fails. The main text therefore measures Y from the end of the measurement window (a landmark design), treating early deaths as a competing event.

Assumption 1 (censoring ignorability). Let V° denote the underlying analyte level that the test would assay, a component of $Z^\star$ that exists whether or not a test is performed. Assumption 1 states that $V^\circ \perp\!\!\!\perp C \mid S^\star, Z$: once ordering intent and the representation are known, censoring carries no further information about the analyte level. It is stated for V° rather than for the recorded value V, because $C = 1$ implies $S = 0$, where V is undefined (Assumption 3). No result in this paper uses Assumption 1; it is the condition under which inverse-probability-weighted audits of ordering

(Section S2.4) are interpretable.

Assumption 6. Assumption 6 is stated as two conditions, 6a (exclusion) and 6b (representation sufficiency), because these are exactly the two properties the proof of Theorem 2 uses. A single condition such as $Y \perp\!\!\!\perp (Z^\star, S, E) \mid Z$ would imply $Y \perp\!\!\!\perp E \mid Z$ directly, so Theorem 2 would follow from it alone and Assumption 4 would play no role; nor would it imply the environment-free posterior that the proof requires.

Two remarks clarify the structure. First, Assumption 4 is not background truth; it is the hypothesis whose identifiability is at issue. Second, we deliberately do not state Assumption 6 as the single condition $Y \perp\!\!\!\perp E \mid Z$, because that condition is the observable implication to be tested (main Section 4.2); assuming it would make the test circular. In the directed acyclic graph (DAG) $E \to Z^\star \to Z$ (Figure S1a), Assumption 6b is not implied by the graph: Z is a child of $Z^\star$, not a separator of E from $Z^\star$. It holds when Z determines $Z^\star$ or when parameters happen to cancel, and main Section 4.3 shows that under Assumptions 5 and 7 such cancellation is non-generic. Assumptions 5 and 7 are not worst-case contingencies; they describe essentially every multisite EHR cohort and every deployed learned representation.

S1.3 Model class

Definition 1 (model class).

The class $\mathcal{M}$ consists of structural causal models M with a latent state space $\mathcal{Z}_M^\star$, mechanisms $E := f_E(U_E)$, $Z^\star := f_{Z^\star}(E, U_{Z^\star})$, $Z := f_Z(Z^\star, U_Z)$, and $Y := f_Y(Z^\star, E, U_Y)$, and mutually independent exogenous variables U. The mechanisms induce the Markov kernels $p_M(e)$, $p_M(z^\star \mid e)$ (case mix), $p_M(z \mid z^\star)$ (representation channel), and $p_M(y \mid z^\star, e)$ (outcome kernel, which we call the outcome mechanism), and through them the observed distribution

$$p_M(y, z, e) = \sum_{z^\star \in \mathcal{Z}_M^\star} p_M(y \mid z^\star, e)\, p_M(z \mid z^\star)\, p_M(z^\star \mid e)\, p_M(e) \tag{S3}$$

(with the sum replaced by an integral for continuous $Z^\star$). The latent state space is not fixed in advance, because $Z^\star$ is never observed and its dimension cannot be learned from data.

The factorization in Eq. (S3) builds in Assumption 6a, since $Y \perp\!\!\!\perp Z \mid (Z^\star, E)$, and it makes the representation channel identical across sites. The second feature is favorable to identification: in reality Z is computed from a record whose observation process differs by site (main Figure 1), so the channel may itself depend on E. Enlarging $\mathcal{M}$ to allow this can only enlarge the set of observationally equivalent models, so non-identifiability within $\mathcal{M}$ implies non-identifiability in any larger class that contains it.

S2 Extended review of related work

The literatures reviewed below each supply one ingredient of our argument: an empirical premise (Sections S2.1 to S2.3), a formal toolkit (Sections S2.4 to S2.6), or a gap that the formal results fill (Section S2.7).

S2.1 Selective measurement and informative missingness in EHR data

The observation that EHR data are shaped by clinical decision-making, and not only by disease, has a substantial empirical literature. Agniel, Kohane, and Weber showed that the presence and timing of an order, independent of its result, carry prognostic information for most common laboratory tests [2]. Pivovarov and colleagues showed that laboratory measurement patterns, such as measurement frequency over time, capture the clinical context of a test, and that aggregating values without separating these patterns can bias EHR analyses [3]. Groenwold showed that missingness in EHR systems is informative because it reflects clinical judgment, and described a "curse of knowing": once clinicians learn that a variable is predictive, their measurement behavior, and with it the informativeness of missingness, can change [4]. Phelan, Bhavsar, and Goldstein described informed presence bias, in which the degree of interaction with a health system shapes which variables are observed at all [5], with conditioning on the number of encounters proposed as one mitigation, itself open to collider bias [6], and Sauer and colleagues catalogued the analytic pitfalls that follow when these features of routinely collected data are ignored [7]. Documentation behavior is itself a clinical signal: the CONCERN early warning system predicts deterioration from real-time patterns of nursing surveillance documentation and was evaluated in a multisite cluster-randomized trial [8]. Together, this literature establishes the empirical premise of our formal results: selective measurement is pervasive, clinically meaningful, and not reducible to random missingness.

S2.2 Representation learning and dataset shift in clinical AI

Embedding methods that integrate sparse, heterogeneous EHR data can improve downstream prediction in some settings [9], but compression does not separate the origins of the compressed signal. Finlayson and colleagues documented how dataset shift in clinical AI arises not only from differences in patient populations but also from changes in clinical practice and local environment [10]; such shift can be invisible to internal validation when training and evaluation data share the same observation regime. Subbaswamy and Saria connected causal structure to shift stability, arguing that a model's robustness in a new environment depends on which causal mechanisms it relies on, not merely on its accuracy in the training distribution [11]. Our results sharpen this line of work for representations built on selectively measured variables: which mechanism a model relies on cannot, in general, be distinguished from an environment-specific artifact using observational data alone.

S2.3 Care-process proxies and algorithmic inequity

A closely related literature documents the consequences of mistaking a health-system response for patient need. Obermeyer and colleagues showed that a population-health algorithm produced racial bias because it used health care spending, a proxy shaped by unequal access, as a stand-in for illness burden [12]. Vyas, Eisenstein, and Jones described the broader danger of encoding social and institutional patterns as biological correction factors in clinical algorithms [13], and Gianfrancesco and colleagues described how missing data and related features of EHR-based machine learning can produce or amplify bias [14]. Our results extend the reach of this literature beyond explicit correction factors: any representation built from selectively measured clinical data is exposed to

the same structural risk, whether or not a protected attribute is involved, because the selection mechanism is itself a channel through which unequal care can enter the representation.

S2.4 Causal inference under selection: identification and its limits

Selective observation has a long formal history. Rubin's framework for missing data [1] distinguishes missingness that is ignorable for likelihood-based inference (missing at random) from missingness that depends on the unobserved values themselves; the EHR ordering process studied here is of the second kind. Cole and Hernán describe inverse-probability weighting as a correction for confounding and selection (for example, censoring) by measured covariates [15]. The same weighting can be used to audit how sensitive a learned representation's apparent performance is to the ordering process, as a check on residual ordering bias rather than a reconstruction of unmeasured values. Bareinboim and Pearl formalized transportability and the data-fusion problem, characterizing when a causal relationship estimated in one environment can be carried to another [16], building on the structural causal model framework [17]. Manski's theory of partial identification shows that when point identification fails, informative bounds can often still be derived under weaker assumptions [18, 19]. This framework bears directly on what remains available once the point-identification result of this paper is accepted.

S2.5 Partial information decomposition

Williams and Beer introduced PID to split the information that two sources jointly carry about a target into redundant, unique, and synergistic components, defined through a small set of axioms rather than a single canonical formula [20]. Subsequent work, including Bertschinger and colleagues [21] and Griffith and Koch [22], proposed alternative measures that satisfy stronger or different properties, reflecting an unresolved debate about which measure suits a given application. Our first formal result concerns a problem that is logically prior to that debate: for EHR value and ordering channels, the joint distribution the decomposition requires is not defined without an encoding choice, and once a sentinel encoding is chosen, the decomposition is determined by it for every measure satisfying the axioms. Choi and colleagues correct PID estimates for missing modalities by inverse-probability weighting under missingness at random [23]; their correction targets values that exist but were not observed, whereas Theorem 1 concerns values that do not exist, for which no complete-data distribution is available to recover.

S2.6 Invariant prediction and domain generalization

Peters, Bühlmann, and Meinshausen proposed invariant causal prediction, which uses the stability of a predictive relationship across heterogeneous environments as evidence that the relationship is causal rather than merely associational; its guarantees hold under stated invariance assumptions, and how much of the causal structure is recovered depends on the richness of the environmental variation [24]. Invariant risk minimization [25] pursues the same intuition with a different estimator, and anchor regression [26] uses heterogeneity across environments to obtain predictors that are robust to bounded distribution shifts. Zhang and colleagues analyzed domain adaptation under target and conditional shift; they recover a changed conditional distribution from unlabeled target-domain data only after restricting how it may change (for example, to a location-scale transformation),

so identification rests on structural assumptions rather than on the data alone [27]. Our central theorem is an analogue of this type of result for the specific structure of EHR case mix and lossy learned representations, proved directly for that setting rather than inherited by citation.

S2.7 The gap this paper addresses

Even within one health system, similar observed EHR states can arise from different underlying physiology, and across systems, differences in measurement, test ordering, documentation, and practice add further variation to what is observed. An observed cross-site difference in a learned representation's relationship to outcomes therefore need not reflect a difference in biology; it may reflect who the sites' patients are, or what the sites do to them. Prior work has documented this ambiguity empirically without a formal identification claim [2, 3, 5], or has relied on general non-identifiability intuitions from causal inference and domain adaptation without a construction specific to selective measurement and learned representations [24, 27]. Closely related non-identifiability and identification results exist for mismeasured mediators, proxy variables, and latent-class models [28–33] (main Section 2); to our knowledge, they have not been applied to cross-site comparison of learned EHR representations under selective measurement. We provide (i) a proof that no division of the information carried by value and ordering channels is identified from a single dataset, with PID as a special case in which a sentinel encoding fixes the decomposition; (ii) an explicit construction of two structural models, differing in whether biological invariance holds, that generate the same observed distribution of outcome, representation, and environment, extended to the full observed record including ordering indicators and laboratory values; (iii) an exact characterization of when the construction applies for binary physiology and a genericity result for finite state spaces; (iv) a worked numerical instance that can be verified by hand; and (v) conditions under which routinely measured tests restore identification of biological invariance.

S3 Single-dataset decomposition: full statements and proofs

Any division of the information about Y between a laboratory's value and the decision to order it requires a joint distribution of the outcome, the value, and the ordering indicator on a common probability space. Two candidates exist for the value. The analyte level V°, the value the assay would return (main Section 3.2), is a random variable on all of Ω, but it is observed only on $\{S = 1\}$. The recorded value V is not a random variable on all of Ω under Assumption 3, so the object $p(y, v, s)$ does not exist until the analyst supplies one, most commonly by the sentinel encoding of Eq. (S2). Part (i) of the following theorem concerns the first candidate, parts (ii) to (iv) the second.

Theorem 1 (a single dataset cannot separate value from ordering).

Let Assumption 2 hold.

(i) Analyte level. *Let Y and V° be discrete, suppose $P(S = 0) > 0$, $H(Y \mid S = 0) > 0$, and $p(y, v^\circ \mid S = 1) > 0$ for every (y, v°), and let the model class place no restriction on $p(v^\circ \mid y, S = 0)$ beyond Assumptions 2 and 6a and the observed distribution of $(Y, S, \tilde{V})$. Then that distribution determines $p(y, s)$ and $p(y, v^\circ \mid S = 1)$; Assumptions 2 and 6a require $p(v^\circ \mid y, S = 0) > 0$ for every (y, v°) and restrict the joint law of (Y, V°, S) no further; and $I(Y; V^\circ \mid S)$ takes every value in the interval from $P(S{=}1)\, I(Y; V^\circ \mid S{=}1)$, included, to $P(S{=}1)\, I(Y; V^\circ \mid S{=}1) + P(S{=}0)\, H(Y \mid S{=}0)$,*

excluded, over the compatible laws, provided V° can take at least as many values as Y does on $\{S = 0\}$ (values arbitrarily close to the upper end are then attained). Consequently $I(Y; V^\circ, S)$ is not identified, and for any bivariate decomposition of $I(Y; V^\circ, S)$ into redundant, unique, and synergistic terms, at least one term is not identified.

For parts (ii) to (iv), let $\varnothing$ lie outside the range of V and let $\tilde{V}$ be defined by Eq. (S2).

(ii) $S = \mathbf{1}[\tilde{V} \neq \varnothing]$ almost surely. Consequently $H(S \mid \tilde{V}) = 0$ and

$$I(\tilde{V}; S) = H(S) > 0, \tag{S4}$$

whatever the joint law of $(Z^\star, S^\star, C, V)$; in particular, Eq. (S4) holds even if ordering is independent of physiology.

(iii) Let Red *be any redundancy measure that satisfies the Williams–Beer axioms [20] (symmetry, self-redundancy, and monotonicity including its equality condition, as stated explicitly by Bertschinger and colleagues [21]) and is invariant under injective re-encoding of a source. Then, for every outcome Y,*

$$\begin{aligned} \mathrm{Red}(Y; \tilde{V}, S) &= I(Y; S), \\ \mathrm{Unq}(Y; S \setminus \tilde{V}) &= 0, \\ \mathrm{Syn}(Y; \tilde{V}, S) &= 0, \\ \mathrm{Unq}(Y; \tilde{V} \setminus S) &= I(Y; \tilde{V}) - I(Y; S). \end{aligned} \tag{S5}$$

(iv) If, in addition, the recorded value carries no outcome information beyond the fact of ordering ($Y \perp\!\!\!\perp \tilde{V} \mid S$) and $I(Y; S) > 0$, then the decomposition reports $\mathrm{Red}(Y; \tilde{V}, S) = I(Y; S) > 0$ *and* $\mathrm{Unq}(Y; \tilde{V} \setminus S) = 0$.

Proof. (i) The observed distribution of $(Y, S, \tilde{V})$ is $p(y, s)$ together with $p(y, v \mid S = 1)$, and $V = V^\circ$ on $\{S = 1\}$; on $\{S = 0\}$ only $p(y \mid S = 0)$ is observed. For a conditional law $q(v^\circ \mid y)$, let $p_q(y, v^\circ, s) := p(y, v^\circ, S{=}1)\, \mathbf{1}[s = 1] + p(y, S{=}0)\, q(v^\circ \mid y)\, \mathbf{1}[s = 0]$, which reproduces the observed distribution.

Which q are compatible. Under Assumption 6a, $Y \perp\!\!\!\perp S \mid (Z^\star, E)$, and under Assumption 2, $0 < P(S = 1 \mid z^\star, e) < 1$. For any $z^\star$ with $V^\circ = v^\circ$, $P(Y{=}y, S{=}s \mid z^\star, e) = p(y \mid z^\star, e)\, P(S{=}s \mid z^\star, e)$, so $p(y, v^\circ, S{=}0) > 0$ if and only if $p(y, v^\circ, S{=}1) > 0$. With $p(y, v^\circ \mid S{=}1) > 0$ everywhere, a compatible q must therefore be positive everywhere. Conversely, every positive q is compatible: take $Z^\star = (V^\circ, W)$ with W taking values in the range of Y, $p(z^\star) := p_q(Y{=}w, V^\circ{=}v^\circ)$, $Y := W$, no censoring, and $P(S^\star{=}1 \mid z^\star) := p_q(S{=}1 \mid Y{=}w, V^\circ{=}v^\circ) \in (0, 1)$. This model generates p_q, satisfies Assumption 2, and satisfies Assumption 6a because Y is a function of $Z^\star$.

The range of the conditional information. Under p_q, $I(Y; V^\circ \mid S) = P(S{=}1)\, I(Y; V^\circ \mid S{=}1) + P(S{=}0)\, I_q(Y; V^\circ \mid S{=}0)$. A positive q free of y gives $I_q(Y; V^\circ \mid S{=}0) = 0$. If V° can take at least as many values as Y does on $\{S = 0\}$, let g be injective and $q_\eta(\cdot \mid y) := (1{-}\eta)\, \delta_{g(y)} + \eta\, u$, with u uniform and $\eta \in (0, 1)$; as $\eta \to 0$, $I_{q_\eta}(Y; V^\circ \mid S{=}0) \to H(Y \mid S{=}0)$, which no compatible q attains because attaining it would make Y a function of V° on $\{S = 0\}$ and so require zeros in q. The set of positive q is convex and $q \mapsto I_q(Y; V^\circ \mid S{=}0)$ is continuous on it, so every value in between is attained. Since $I(Y; S)$ is fixed by the data and $H(Y \mid S{=}0) > 0$, $I(Y; V^\circ, S) = I(Y; S) + I(Y; V^\circ \mid S)$ takes more than one value. The terms of a bivariate decomposition sum to $I(Y; V^\circ, S)$, so if all of them were identified, their sum would be too.

(ii) On $\{S = 1\}$, $\tilde{V} = V \neq \varnothing$ because $\varnothing$ is outside the range of V; on $\{S = 0\}$, $\tilde{V} = \varnothing$. Hence $S = g(\tilde{V})$ with $g(v) := \mathbf{1}[v \neq \varnothing]$, so $H(S \mid \tilde{V}) = 0$ and $I(\tilde{V}; S) = H(S) - H(S \mid \tilde{V}) = H(S)$. By Assumption 2, $0 < P(S = 1) < 1$, so $H(S) > 0$. No property of $(Z^\star, S^\star, C, V)$ was used. Because $I(\tilde{V}; S) \leq H(S)$ always holds for discrete S, the sentinel attains the largest possible value.

(iii) Treat each source as a collection of variables. Since $\{S\} \subseteq \{\tilde{V}, S\}$, the equality condition of the monotonicity axiom gives $\mathrm{Red}(Y; \{S\}, \{\tilde{V}, S\}) = \mathrm{Red}(Y; \{S\})$, which equals $I(Y; S)$ by self-redundancy. Because $S = g(\tilde{V})$, the map $\tilde{V} \mapsto (\tilde{V}, g(\tilde{V}))$ is injective, so re-encoding invariance gives $\mathrm{Red}(Y; \{S\}, \{\tilde{V}\}) = I(Y; S)$, and symmetry gives the first identity in Eq. (S5). The bivariate decomposition satisfies $I(Y; S) = \mathrm{Red} + \mathrm{Unq}_S$, $I(Y; \tilde{V}) = \mathrm{Red} + \mathrm{Unq}_{\tilde{V}}$, and $I(Y; \tilde{V}, S) = \mathrm{Red} + \mathrm{Unq}_{\tilde{V}} + \mathrm{Unq}_S + \mathrm{Syn}$. The first yields $\mathrm{Unq}_S = 0$, the second yields $\mathrm{Unq}_{\tilde{V}} = I(Y; \tilde{V}) - I(Y; S)$, which is non-negative by the data-processing inequality, and, since $I(Y; \tilde{V}, S) = I(Y; \tilde{V})$ when S is a function of $\tilde{V}$, the third yields $\mathrm{Syn} = 0$.

(iv) By the chain rule, $I(Y; \tilde{V}) = I(Y; \tilde{V}, S) = I(Y; S) + I(Y; \tilde{V} \mid S) = I(Y; S)$ when $Y \perp\!\!\!\perp \tilde{V} \mid S$; substitute in Eq. (S5). □

Interpretation. Part (i) is a recoverability result in the sense of Mohan and Pearl [34, 35]; compare their Theorems 2 and 3, under which the joint distribution, or a conditional distribution of a variable, is not recoverable when the variable is adjacent to its own missingness mechanism. Here the analyte reaches S through ordering intent, which is latent and unrestricted, so marginalizing over it yields that adjacency. Their m-graphs also assume that no variable is a child of a missingness indicator, whereas here V and Z are children of S; the argument of part (i) stands on its own, and the link to m-graphs is interpretive. Part (i) needs no axioms about how information is decomposed, and it is removed only by assumptions under which $p(v^\circ \mid y, S = 0)$ is recoverable (for example $V^\circ \perp\!\!\!\perp S \mid Y$), or by measuring every patient (main Section 4.5). Part (ii) shows that the sentinel creates an information channel about ordering, not by revealing anything about physiology, but because the code for "not measured" is itself a record of the ordering decision. Part (iii) shows the consequence for PID: every admissible redundancy measure attributes all of the ordering indicator's predictive content to information "shared" with the value channel, assigns the ordering indicator no unique information, and assigns no synergy. The decomposition is therefore fixed by the encoding, and the choice among redundancy measures [20–22] does not matter. Part (iv) makes the artifact explicit: when only the decision to order predicts the outcome, the decomposition reports that the value and ordering channels carry identical information. The bound $\mathrm{Red}(Y; \tilde{V}, S) \leq \min\{I(Y; \tilde{V}), I(Y; S)\}$, which follows from monotonicity and self-redundancy, is attained at $I(Y; S)$.

Re-encoding invariance of the cited measures. Part (iii) assumes invariance under injective re-encoding of a source, a property beyond the Williams–Beer axioms. It holds for the three measures cited, because each depends on a source only through the joint law of that source with Y, and this law is unchanged, up to relabeling of the source alphabet, by an injective map. For these measures the conclusion of part (iii) can also be checked directly. For $I_{\min}$ [20], the specific information $I(Y{=}y; A) = D_{\mathrm{KL}}(p(a \mid y) \,\|\, p(a))$ cannot increase under the map $\tilde{V} \mapsto S = g(\tilde{V})$, by the data-processing inequality for relative entropy, so the minimum over the two sources is attained at S for every y and $I_{\min} = I(Y; S)$. For the measure of Bertschinger and colleagues [21], this follows from their characterization of vanishing unique information, which, through Blackwell's theorem, implies that a source that is a garbling of another carries no unique information:

$\mathrm{Unq}(Y; S \setminus \tilde{V}) = \min_{Q \in \Delta_P} I_Q(Y; S \mid \tilde{V})$, where Δ_P is the set of joint laws with the same (Y, S) and $(Y, \tilde{V})$ marginals as P; the minimum is 0, attained at $Q = P$. For the union information of Griffith and Koch [22], $\min_{Q \in \Delta_P} I_Q(Y; \tilde{V}, S) \geq I(Y; \tilde{V})$ because every $Q \in \Delta_P$ has the $(Y, \tilde{V})$ marginal of P, and $Q = P$ attains it; the implied redundancy is $I(Y; S) + I(Y; \tilde{V}) - I(Y; \tilde{V}) = I(Y; S)$.

A numerical instance of part (i). Let Y, V°, and S be binary with $P(S = 1) = 0.5$, $P(Y{=}y, V^\circ{=}v \mid S{=}1) = 0.30, 0.10, 0.05, 0.55$ for $(y, v) = (0,0), (0,1), (1,0), (1,1)$, and $P(Y{=}1 \mid S{=}0) = 0.3$. Completion A takes V° independent of Y on $\{S = 0\}$ with $P(V^\circ{=}1 \mid S{=}0) = 0.5$; completion B takes $P(V^\circ{=}y \mid Y{=}y, S{=}0) = 0.95$. Both are positive, so both are compatible with Assumptions 2 and 6a, and both reproduce the observed distribution exactly, but $I(Y; V^\circ, S) = 0.247$ bits under A and 0.556 bits under B, and $I(Y; V^\circ) = 0.103$ and 0.536 bits. For the redundancy measures of Williams and Beer and of Bertschinger and colleagues, the redundant term equals $I(Y; S) = 0.067$ bits in both completions, so the unique term of the value (0.036 against 0.469 bits) and the synergistic term (0.144 against 0.021 bits) carry the difference. The corresponding interval for $I(Y; V^\circ, S)$ runs here from 0.247 to 0.688 bits, the upper end approached as the completion concentrates on $V^\circ = Y$.

Clinical example. Suppose lactate is ordered in the emergency department mainly when a clinician suspects sepsis, and unmeasured lactate is stored as -1 (main Figure 2b). Any model or decomposition applied to $\tilde{V}$ receives, in the -1 code, a perfect record of which patients the clinician did not suspect. If clinician suspicion predicts mortality, PID will report that "lactate" and "ordering" share that information, even in the extreme case in which the measured lactate values themselves add nothing.

Sentinel encoding is the most common construction but not the only one. The next result shows that every alternative faces the same problem in a different form.

Corollary 1 (no encoding-neutral resolution).

Under Assumptions 2 and 3, any decomposition of the information about Y into "value" and "ordering" contributions must adopt one of the following constructions, and none recovers the separate informational contributions of physiology and care process:

(a) Restriction to measured patients. *On $\{S = 1\}$, S is constant, so $I(Y; S \mid S = 1) = 0$ and, by monotonicity, self-redundancy, and nonnegativity of the redundancy measure, the redundant and unique ordering components are zero. The ordering channel is not decomposed; it is removed.*

(b) Extension of the domain by a fill-in rule. *Let $\tilde{V} = V$ on $\{S = 1\}$ and $\tilde{V} = W$ on $\{S = 0\}$, for any random element W (a sentinel, a constant, or an imputation). Then $I(\tilde{V}; S) = 0$ if and only if the law of W given $S = 0$ equals the law of V given $S = 1$; otherwise the information $I(\tilde{V}; S) > 0$ is a property of the fill-in rule. The sentinel is the extreme case, $I(\tilde{V}; S) = H(S)$ (Theorem 1).*

(c) Positing a value for unmeasured patients. *Treating a "value had it been measured" as a random variable on $\{S = 0\}$ contradicts Assumption 3. If Assumption 3 is relaxed by using the analyte level V°, its law on $\{S = 0\}$ is not identified (Theorem 1(i)), so the decomposition depends on an unidentified distribution; choosing that distribution to equal the law of V given $S = 1$, which makes the fill-in in (b) information-neutral, amounts to assuming that unmeasured patients resemble measured ones, the premise that selective measurement contradicts (Section S2.1).*

Proof. (a) is immediate. For (b), because $0 < P(S = 1) < 1$, $\tilde{V} \perp\!\!\!\perp S$ holds if and only if the conditional laws of $\tilde{V}$ given $S = 1$ and given $S = 0$ coincide, and these are the law of V given $S = 1$

and the law of W given $S = 0$. (c) follows from Assumption 3 and from the fact that no observation is made on $\{S = 0\}$.

□

Corollary 1 makes precise the informal claim that "any resolution manufactures information." The precise statement is that each resolution either removes the ordering channel, makes its informational content a property of the encoding, or makes the decomposition depend on an assumed distribution for values that were never generated. This conclusion concerns the well-definedness of the object being decomposed. It does not adjudicate among redundancy measures in settings where the domain problem does not arise.

If a single dataset cannot separate the channels, a natural next step is to compare the same relationship across independent environments, on the premise that biology does not depend on institutional context while practice does.

S4 Cross-site invariance and non-identifiability: additional proofs and results

Proposition 1 (passing the test does not establish biological invariance).

Let Y be binary and let P_{obs} be any distribution of (Y, Z, E) with finite $\mathcal{Z}$ that satisfies main Eq. (6), with $\mathcal{Z}_{\text{sh}}$ non-empty and $r_e(z) := P_{\text{obs}}(Y = 1 \mid z, e) \in (0, 1)$ wherever $P_{\text{obs}}(z \mid e) > 0$. Then there is a model $M_3 \in \mathcal{M}$ that generates P_{obs}, satisfies Assumptions 6a, 6b, and 7, and violates Assumption 4. It also satisfies Assumption 5 whenever $P_{\text{obs}}(z \mid e)$ varies with e.

Proof. Let $Z^\star_{M_3} := (Z, U)$ with $U \in \{0, 1\}$, $P(U = 1) = 1/2$, and U independent of (Z, E); let the representation channel be the projection $(z, u) \mapsto z$ and the case mix $p_{M_3}((z, u) \mid e) := P_{\text{obs}}(z \mid e)/2$. Choose constants δ_e with $|\delta_e| < \min_{z,e} \min\{r_e(z), 1 - r_e(z)\}$, not all equal, and set $P_{M_3}(Y = 1 \mid (z, u), e) := r_e(z) + \delta_e(2u - 1)$ wherever $P_{\text{obs}}(z \mid e) > 0$. Averaging over U gives $P_{M_3}(Y = 1 \mid z, e) = r_e(z)$, so M_3 reproduces P_{obs}. For $z \in \mathcal{Z}_{\text{sh}}$, main Eq. (6) makes $r_e(z)$ the same at every site, so the mechanism at (z, u) depends on e through δ_e, and Assumption 4 fails. Since Z is a function of $Z^\star_{M_3}$, Assumption 6a holds; the posterior $p_{M_3}((z', u) \mid z, e) = \mathbf{1}[z' = z]/2$ does not depend on e, so Assumption 6b holds; and U is not recoverable from Z, so Assumption 7 holds.

□

Remark 1 (biological invariance alone has no testable implication).

Over $\mathcal{M}$, every distribution of (Y, Z, E) on finite alphabets is generated by some model satisfying Assumption 4: take $Z^\star := (Z, Y)$, case mix $p(z^\star \mid e) := P_{\text{obs}}(z, y \mid e)$, the projection onto the first coordinate as the representation channel, and the site-invariant, deterministic mechanism $p(y \mid z^\star) := \mathbf{1}[y = z^\star_2]$. Assumption 4 therefore restricts the observed distribution only in conjunction with Assumption 6, which is why Theorem 2 requires both.

In M_3, the sites treat identical patients differently, but the differences cancel within every representation cell. A passed invariance test is therefore a statement about a margin of the joint distribution, not about the mechanism beneath it. The natural next question is the converse: when

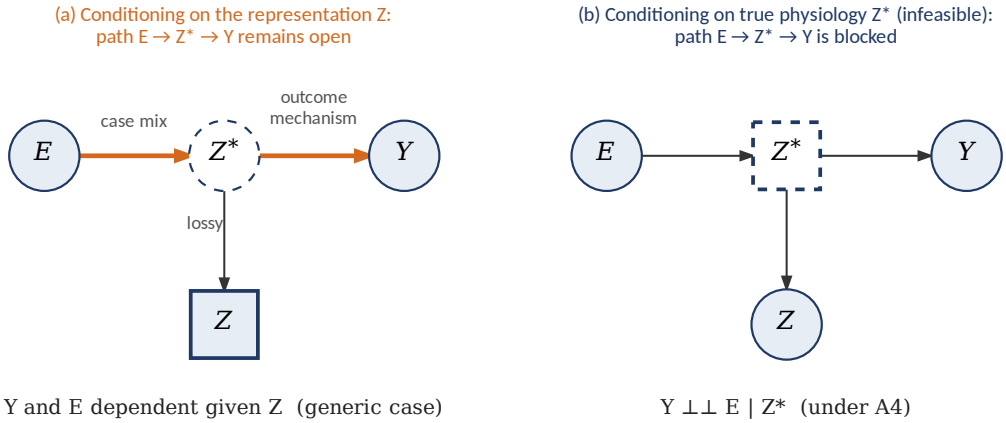


Figure S1: **Representation loss and case-mix heterogeneity.** Squares denote conditioned variables. (a) Conditioning on the learned representation Z leaves the path $E \to Z^\star \to Y$ open (orange), because Z is a lossy child of $Z^\star$ and not a node on the path; different sites therefore contribute different mixtures of true physiology to the same cell $Z = z$, and the outcome rate in that cell differs across sites even though the mechanism $Z^\star \to Y$ is shared. (b) Conditioning on $Z^\star$ would block the path, and under Assumption 4 would make Y independent of E; this is the comparison researchers would like to make but cannot. The reader should take away that conditioning on a representation is not conditioning on physiology.

the test fails, can the failure at least be attributed to biology? Main Section 4.3 shows that it cannot.

Expanding $p_{M_1}(y \mid z, e)$ under Assumption 4 and Bayes' rule gives

$$\begin{aligned} p_{M_1}(y \mid z, e) &= \sum_{z^\star} p_{M_1}(y \mid z^\star)\, p_{M_1}(z^\star \mid z, e), \\ p_{M_1}(z^\star \mid z, e) &= \frac{p_{M_1}(z \mid z^\star)\, p_{M_1}(z^\star \mid e)}{\sum_{z'} p_{M_1}(z \mid z')\, p_{M_1}(z' \mid e)}. \end{aligned} \tag{S6}$$

The mechanism $p_{M_1}(y \mid z^\star)$ is the same at every site. All site dependence enters through the posterior, and it enters there only through the case-mix prior $p_{M_1}(z^\star \mid e)$. The channel then decides whether that prior matters: if Z pins down $Z^\star$, the posterior is a point mass whatever the prior. Figure S1 shows the same point graphically. Conditioning on Z does not block the path $E \to Z^\star \to Y$, because Z is a descendant of $Z^\star$ rather than a node on the path; only conditioning on $Z^\star$ itself, which is infeasible, would block it [17].

Proposition 2 (exact conditions for binary physiology).

Let M_1 satisfy Assumption 4 with $\mathcal{Z}^\star = \{0, 1\}$ and write $\mu_k := p_{M_1}(Y = 1 \mid Z^\star = k)$, $q_e := p_{M_1}(Z^\star = 1 \mid e)$, $a_z := p_{M_1}(Z = z \mid Z^\star = 1)$, and $b_z := p_{M_1}(Z = z \mid Z^\star = 0)$. For z with

$p_{M_1}(z \mid e) > 0$ *and* $p_{M_1}(z \mid e') > 0$,

$$p_{M_1}(Y{=}1 \mid z, e) = \mu_0 + (\mu_1 - \mu_0)\,\pi_e(z), \qquad \pi_e(z) := \frac{a_z q_e}{a_z q_e + b_z(1 - q_e)}, \tag{S7}$$

so that the cross-site gap is

$$\Delta_{e,e'}(z) := p_{M_1}(Y{=}1 \mid z, e) - p_{M_1}(Y{=}1 \mid z, e') = (\mu_1 - \mu_0)\,[\pi_e(z) - \pi_{e'}(z)]. \tag{S8}$$

The gap is non-zero if and only if all three of the following hold: (a) $\mu_1 \neq \mu_0$ (physiology affects the outcome); (b) $q_e \neq q_{e'}$ (case mix differs, Assumption 5); and (c) $a_z > 0$ and $b_z > 0$ (both physiological states can produce the representation value z, so the representation is lossy at z, Assumption 7).

Proof. Eq. (S7) is Eq. (S6) with binary $Z^\star$, and Eq. (S8) follows. If $a_z = 0$ or $b_z = 0$, then $\pi_e(z) \in \{0, 1\}$ takes the same value for every e. If $a_z, b_z > 0$, then $d\pi/dq = a_z b_z/[a_z q + b_z(1-q)]^2 > 0$, so π is strictly increasing in q on $[0, 1]$ and $\pi_e(z) = \pi_{e'}(z)$ if and only if $q_e = q_{e'}$.

□

For binary physiology, then, the conditions under which the construction applies are exactly the three conditions summarized in the main Abstract: case-mix heterogeneity, a lossy representation, and physiology that matters for the outcome. For larger state spaces, exact cancellations become possible, and the appropriate statement is generic.

Proposition 3 (genericity for finite state spaces).

Let $\mathcal{Z}^\star$, $\mathcal{Z}$, $\mathcal{Y}$, and $\mathcal{E}$ be finite, each with at least two elements, and fix $p(e) > 0$. Parametrize the models in $\mathcal{M}$ that satisfy Assumption 4 by $\theta = (\{p(z^\star \mid e)\}, \{p(z \mid z^\star)\}, \{p(y \mid z^\star)\})$, ranging over a product of probability simplices Θ. The set of θ for which $p_\theta(y \mid z, e)$ is invariant in e for all y and all $z \in \mathcal{Z}_{\mathrm{sh}}$ has Lebesgue measure zero in Θ.

Proof. Fix $y_0 \in \mathcal{Y}$, $z_0 \in \mathcal{Z}$, and $e_0 \neq e_1$, and let $D_e(\theta) := p_\theta(z_0 \mid e)$ and $N_e(\theta) := p_\theta(y_0, z_0 \mid e) = \sum_{z^\star} p(y_0 \mid z^\star) p(z_0 \mid z^\star) p(z^\star \mid e)$, both polynomials in θ. Wherever $D_{e_0}, D_{e_1} > 0$, invariance at (y_0, z_0) requires $F(\theta) := N_{e_1}(\theta) D_{e_0}(\theta) - N_{e_0}(\theta) D_{e_1}(\theta) = 0$. Embedding the example of Section S5 (placing all mass on two elements of each alphabet) gives $D_{e_0} = 0.34$, $D_{e_1} = 0.66$, $N_{e_0} = 0.23$, $N_{e_1} = 0.51$ and hence $F = 0.51 \times 0.34 - 0.23 \times 0.66 = 0.0216 \neq 0$, so F is not identically zero on the affine hull of Θ. The zero set of a polynomial that is not identically zero has Lebesgue measure zero [36], and the same applies to the zero set of each D_e, $e \in \mathcal{E}$, which are also non-zero polynomials. If $z_0 \notin \mathcal{Z}_{\mathrm{sh}}$, some D_e vanishes; otherwise invariance requires $F = 0$. The invariant set is therefore contained in $\{F = 0\} \cup \bigcup_e \{D_e = 0\}$, a finite union of null sets.

□

Proposition 3 makes precise the sense in which failure of Assumption 6b under Assumptions 5 and 7 is "generic": parameter values at which case-mix differences cancel exactly within every representation cell form a set of measure zero. This parallels the standard result that parameter values violating faithfulness have measure zero (for discrete networks [37]; for linear-Gaussian models [38]). It is a statement about what cannot be relied on, not a claim that cancellation is logically impossible.

Remark 2 (the equivalent model can satisfy every descriptive assumption).

In Lemma 1 the identity channel makes M_2 lossless, so M_2 violates Assumption 7. This is inessential. Let $Z^\star_{M_2'} := (Z, U)$ with $U \sim \text{Bernoulli}(1/2)$ independent of everything else, use the projection $(z, u) \mapsto z$ as the channel, set $p_{M_2'}((z, u) \mid e) := p_{M_1}(z \mid e)/2$ and $p_{M_2'}(y \mid (z, u), e) := p_{M_1}(y \mid z, e)$. Then $M_2' \sim M_1$; M_2' violates Assumption 4 under the same condition as M_2; it satisfies Assumptions 6a, 6b, and 7; and it satisfies Assumption 5 whenever $p_{M_1}(z \mid e)$ varies with e, which is itself observable. Restricting $\mathcal{M}$ to models that satisfy Assumptions 5 and 7 therefore does not restore identifiability. Nor can the cardinality of $\mathcal{Z}^\star$ be used to exclude M_2, because it is not identified from data.

Theorem 3 is stated for the observed triple (Y, Z, E). In practice the analyst also observes the ordering indicator and the laboratory values, which might seem to provide extra leverage. The following corollary shows that, without further restrictions on how the record depends on physiology, they do not resolve the ambiguity unless they pin down the outcome-relevant physiology.

Corollary 2 (extension to the full observed record).

Let $W := (Z, \tilde{V})$ denote the observed record, which determines S by Theorem 1(ii). Consider models that specify $p(w \mid z^\star, e)$, allowing site-specific ordering and measurement, together with $p(y \mid z^\star, e)$ and $p(z^\star \mid e)$, and that satisfy the extended exclusion $Y \perp\!\!\!\perp W \mid (Z^\star, E)$. Applying Lemma 1 with W in place of Z yields a model M_2 that is observationally equivalent to M_1 on (Y, W, E), and hence on the observed (y, z, s, v, e), and that violates Assumption 4 if and only if $p_{M_1}(y \mid w, e)$ varies with e. Under Assumption 4 in M_1, $p_{M_1}(y \mid w, e) = \sum_{z^\star} p(y \mid z^\star)\, p(z^\star \mid w, e)$ with $p(z^\star \mid w, e) \propto p(w \mid z^\star, e)\, p(z^\star \mid e)$, so site dependence now enters through both case mix and site-specific ordering.

Proof. The proof of Lemma 1 uses only the factorization of Eq. (S3), which holds with W in place of Z and with a site-dependent channel (the channel does not appear in the collapsed model). Here M_1 belongs to the enlarged class with a site-dependent channel $p(w \mid z^\star, e)$, while M_2 has the identity channel and so belongs to $\mathcal{M}$ with W in place of Z.

□

Corollary 2 qualifies the informal claim that V and S "add no information." They can add information: if the full record determined the outcome-relevant component of $Z^\star$, the posterior would degenerate and the ambiguity would disappear. Absent restrictions such as conditionally independent proxies or a bounded latent state space, under which latent-variable identification results can apply, they do not add information *of the right kind* unless they do so, and when ordering practice differs by site they introduce a further source of site dependence into the posterior. This is the formal reason for anchoring invariance tests on protocolized, near-universally ordered measurements (main Section 5.3): such measurements remove the site-specific ordering pathway, although they do not by themselves remove the case-mix pathway of Proposition 2.

S4.1 Routine and multiple measurements: proofs and checks

Proposition 4. The statement and proof are in main Section 4.5. Two remarks. First, the condition that the outcome depends on physiology only through $V = g(Z^\star)$ is the form that sufficiency takes in the model of main Figure 1: there a measured value has no edge to the outcome,

so $Y \perp\!\!\!\perp V \mid (Z^\star, E)$, and if V were a noisy measurement with positive conditional probabilities, requiring also $Y \perp\!\!\!\perp Z^\star \mid (V, E)$ would force $Y \perp\!\!\!\perp (V, Z^\star) \mid E$ by the intersection property, so that physiology would not affect the outcome. Sufficiency is therefore non-trivial only for an exact summary. Second, the proof uses only exclusion and the support condition, so the result holds within the tested subpopulation when ordering depends only on physiology and site, since then $p(y \mid v, e, S{=}1) = h_e(y \mid v)$, and Assumption 2, which holds at each site, guarantees that every state occurs among tested patients at every site where it occurs at all.

Corollary 3. Because $V = Z^\star$, each quantity defined in terms of $Z^\star$ equals the same quantity defined in terms of the observed V: $q_e = P(V{=}1 \mid e)$, $a_z = P(Z{=}z \mid V{=}1)$, $b_z = P(Z{=}z \mid V{=}0)$ (the channel is site-invariant in the model class, so these may be computed at any site or pooled), and $\mu_{k,e} = P(Y{=}1 \mid V{=}k, e)$. Eq. (S8) then gives the case-mix part of a cross-site gap, and the remainder is the mechanism part.

Proposition 5 (two tests and a case-mix difference identify biological invariance).

Let $Z^\star \in \{0,1\}$, with two sites of case mix $q_0 \neq q_1$, both in $(0,1)$. Let binary tests V_1, V_2 be performed for every patient at both sites, with sensitivities $\mathrm{sens}_j = P(V_j{=}1 \mid Z^\star{=}1)$ and specificities $\mathrm{spec}_j = P(V_j{=}0 \mid Z^\star{=}0)$ that are the same at both sites and satisfy $\mathrm{sens}_j + \mathrm{spec}_j > 1$, and let V_1, V_2, and Y be conditionally independent given $(Z^\star, E)$. Then q_0, q_1, the four error rates, and $\mu_{k,e} = P(Y{=}1 \mid Z^\star{=}k, E{=}e)$ are identified from the distribution of (Y, V_1, V_2, E).

Proof. Step 1 (prevalences and error rates). At site e, $P(V_1{=}v_1, V_2{=}v_2 \mid e) = q_e \prod_j P(V_j{=}v_j \mid Z^\star{=}1) + (1 - q_e) \prod_j P(V_j{=}v_j \mid Z^\star{=}0)$. The two sites give six free observed probabilities for six parameters $(q_0, q_1, \mathrm{sens}_1, \mathrm{spec}_1, \mathrm{sens}_2, \mathrm{spec}_2)$, and Hui and Walter [33] show that when the prevalences differ the system has exactly two solutions, which are exchanged by relabeling $Z^\star$ (replacing each prevalence by its complement and exchanging each test's sensitivity with one minus its specificity). Relabeling turns $\mathrm{sens}_j + \mathrm{spec}_j > 1$ into $\mathrm{sens}_j + \mathrm{spec}_j < 1$, so the stated condition selects one solution.

Step 2 (outcome mechanism). By conditional independence, for $v_1 \in \{0,1\}$,

$$P(Y{=}1, V_1{=}v_1 \mid e) = \sum_{k \in \{0,1\}} P(Z^\star{=}k \mid e)\, P(V_1{=}v_1 \mid Z^\star{=}k)\, \mu_{k,e}.$$

The coefficient matrix has rows $\big((1-q_e)\,\mathrm{spec}_1,\ q_e(1-\mathrm{sens}_1)\big)$ and $\big((1-q_e)(1-\mathrm{spec}_1),\ q_e\,\mathrm{sens}_1\big)$ and determinant $q_e(1-q_e)(\mathrm{sens}_1 + \mathrm{spec}_1 - 1) \neq 0$, so $(\mu_{0,e}, \mu_{1,e})$ is the unique solution. □

Numerical check. In 15 randomly generated models (case mix differing by at least 0.15, sensitivities and specificities in $[0.65, 0.97]$, and Assumption 4 holding in about half of them), every parameter vector that reproduced the observed distribution of (Y, V_1, V_2, E) exactly, found from 40 random starts each, agreed with the true one to within 10^{-8} after fixing labels as above, including the site-specific mechanism. The script is included with the code accompanying this article.

As with Proposition 4, the argument uses exclusion rather than routine measurement as such, so it also applies to the subpopulation in which both tests were ordered, provided ordering depends only on physiology and site and the case mix among tested patients, $P(Z^\star{=}1 \mid e, S{=}1)$, still differs between the two sites; site-dependent ordering can equalize it. The conditional independence of the two tests given physiology is an assumption about assays (their errors arise from separate processes),

and it is the assumption under which latent-class estimation of diagnostic accuracy without a gold standard is standard.

S5 The worked example in full

We now build the pair (M_1, M_2) explicitly. The numbers are stylized, chosen for arithmetic transparency rather than clinical realism. For concreteness one may read E as two hospital systems, $Z^\star$ as true sepsis status, Z as a binary high-risk flag derived from a lactate-based embedding, and Y as 30-day mortality; the channel is held fixed across sites to isolate the case-mix mechanism (site-specific ordering would add a further source of site dependence, by Corollary 2). A conditional mortality of 0.8 is higher than is typical in sepsis cohorts; the reading is for orientation only.

Let $E \in \{0, 1\}$ and $Z^\star, Z, Y \in \{0, 1\}$. Table S3 carries out the computation step by step, using Eqs. (S6) and (S7).

Step 1 (case mix). $P(Z^\star = 1 \mid E = 0) = 0.3$ and $P(Z^\star = 1 \mid E = 1) = 0.7$ (Assumption 5).

Step 2 (representation channel). Symmetric and identical at both sites, with 10% error: $P(Z = 1 \mid Z^\star = 1) = 0.9$ and $P(Z = 1 \mid Z^\star = 0) = 0.1$ (Assumption 7: both physiological states produce both representation values).

Step 3 (outcome mechanism in M_1). $P(Y = 1 \mid Z^\star = 0) = 0.2$ and $P(Y = 1 \mid Z^\star = 1) = 0.8$ at both sites (Assumption 4 holds by construction).

Step 4 (representation prevalence). $P(Z = 1 \mid E = 0) = 0.9(0.3) + 0.1(0.7) = 0.34$ and $P(Z = 1 \mid E = 1) = 0.9(0.7) + 0.1(0.3) = 0.66$.

Step 5 (posterior). $P(Z^\star = 1 \mid Z = 1, E = 0) = 0.27/0.34 = 27/34 \approx 0.794$ and $P(Z^\star = 1 \mid Z = 1, E = 1) = 0.63/0.66 = 21/22 \approx 0.955$. Among patients with the same representation, site 1's patients are more often truly in the high-risk state.

Step 6 (observed outcome rates). By Eq. (S7), $P(Y = 1 \mid Z = 1, E = 0) = 0.2 + 0.6(27/34) = 23/34 \approx 0.676$ and $P(Y = 1 \mid Z = 1, E = 1) = 0.2 + 0.6(21/22) = 17/22 \approx 0.773$. The gap is $18/187 \approx 0.096$, or 9.6 percentage points, and by Eq. (S8) it equals $(\mu_1 - \mu_0)(\pi_1 - \pi_0) = 0.6 \times 0.160$. An identical 9.6-point gap arises in the $Z = 0$ cell ($5/22 \approx 0.227$ versus $11/34 \approx 0.324$).

Table S3: Step-by-step computation for M_1 (Assumption 4 holds by construction). Exact fractions are given in Section S5.1.

Quantity	Site $E = 0$	Site $E = 1$	Source
Case mix $P(Z^\star = 1 \mid e)$	0.300	0.700	Step 1 (A5)
Channel $P(Z = 1 \mid Z^\star = 1)$, $P(Z = 1 \mid Z^\star = 0)$	0.9, 0.1	0.9, 0.1	Step 2 (A7)
Mechanism $P(Y = 1 \mid Z^\star = 0)$, $P(Y = 1 \mid Z^\star = 1)$	0.2, 0.8	0.2, 0.8	Step 3 (A4)
$P(Z = 1 \mid e)$	0.340	0.660	Step 4
Posterior $P(Z^\star = 1 \mid Z = 1, e)$	0.794	0.955	Step 5, Eq. (S6)
Posterior $P(Z^\star = 1 \mid Z = 0, e)$	0.045	0.206	Step 5, Eq. (S6)
$P(Y = 1 \mid Z = 1, e)$	0.676	0.773	Step 6, Eq. (S7)

Quantity	Site $E = 0$	Site $E = 1$	Source
$P(Y = 1 \mid Z = 0, e)$	0.227	0.324	Step 6, Eq. (S7)

Step 7 (construct M_2 by Lemma 1). Set $P_{M_2}(Z^\star = 1 \mid E = 0) = 0.34$ and $P_{M_2}(Z^\star = 1 \mid E = 1) = 0.66$ (the observed prevalences of $Z = 1$), let $Z = Z^\star$, and set the mechanism to the observed rates of Step 6: $P_{M_2}(Y = 1 \mid Z^\star = 1, E = 0) = 23/34$, $P_{M_2}(Y = 1 \mid Z^\star = 1, E = 1) = 17/22$, $P_{M_2}(Y = 1 \mid Z^\star = 0, E = 0) = 5/22$, and $P_{M_2}(Y = 1 \mid Z^\star = 0, E = 1) = 11/34$.

Step 8 (observational equivalence). Both models induce the joint distribution $p(y, z \mid e)$ shown in Table S4: at site 0, $P(Y = 1, Z = 1) = 0.23$, $P(Y = 0, Z = 1) = 0.11$, $P(Y = 1, Z = 0) = 0.15$, $P(Y = 0, Z = 0) = 0.51$; at site 1, 0.51, 0.15, 0.11, and 0.23, respectively. With $p(e)$ shared, $p(y, z, e)$ is identical.

Step 9 (the interpretations differ). Transfer a patient whose true state is $Z^\star = 1$ from site 0 to site 1. In M_1 the patient's outcome risk is 0.8 at both sites; the 9.6-point gap arises entirely from who occupies the $Z = 1$ cell. In M_2 the same transfer raises risk from 0.676 to 0.773: the gap is a site effect on the mechanism.

Table S4: M_1 versus M_2: different structures, identical observations.

Quantity	M_1 ("different patients")	M_2 ("different care")
Latent state space	$Z^\star \in \{0, 1\}$, distinct from Z	$Z^\star = Z$
Case mix $P(Z^\star = 1 \mid E = 0)$, $P(Z^\star = 1 \mid E = 1)$	0.30, 0.70	0.34, 0.66
Channel $P(Z = 1 \mid Z^\star = 1)$, $P(Z = 1 \mid Z^\star = 0)$	0.9, 0.1 (lossy)	1, 0 (identity)
Mechanism $P(Y = 1 \mid Z^\star = 1)$ at $E = 0$, $E = 1$	0.800, 0.800	0.676, 0.773
Mechanism $P(Y = 1 \mid Z^\star = 0)$ at $E = 0$, $E = 1$	0.200, 0.200	0.227, 0.324
Assumption 4 (biological invariance)	Holds	Fails
Assumption 6b (representation sufficiency)	Fails	Holds
$P(Y, Z \mid E = 0)$ for $(1,1), (0,1), (1,0), (0,0)$	0.23, 0.11, 0.15, 0.51	0.23, 0.11, 0.15, 0.51
$P(Y, Z \mid E = 1)$ for $(1,1), (0,1), (1,0), (0,0)$	0.51, 0.15, 0.11, 0.23	0.51, 0.15, 0.11, 0.23
Observed $P(Y = 1 \mid Z = 1, e)$ at $E = 0$, $E = 1$	0.676, 0.773	0.676, 0.773
Effect of moving a $Z^\star = 1$ patient from site 0 to site 1	None	+9.6 percentage points

The gap is not a small-sample artifact. With roughly 340 patients per site in the $Z = 1$ stratum,

a two-sided two-proportion test at $\alpha = 0.05$ has about 80% power to detect it, a sample size well within the range of multisite EHR cohorts. Collecting more observational data of this kind at either site sharpens the estimate of the gap but cannot adjudicate between M_1 and M_2.

Sensitivity analysis for the binary case. Equation (S8) turns the example into a calculable sensitivity analysis. Suppose an analyst is willing to posit the representation's error rates. With a symmetric error rate ε (so $a_1 = 1 - \varepsilon$ and $b_1 = \varepsilon$), the case mix is recovered from the observed prevalences as $q_e = [P(Z = 1 \mid e) - \varepsilon]/(1 - 2\varepsilon)$, the posteriors $\pi_e(z)$ follow from Eq. (S7), and, because $|\mu_1 - \mu_0| \leq 1$, Assumption 4 can account for a gap of at most $|\pi_1(z) - \pi_0(z)|$. In the example, this maximum is 0.029 when $\varepsilon = 0.02$, 0.075 when $\varepsilon = 0.05$, and 0.160 when $\varepsilon = 0.10$. The smallest error rate at which case mix alone can explain the observed 9.6-point gap in one cell is $\varepsilon \approx 0.063$ (Table S5). This one-cell bound is a necessary condition, not a sufficient one. Using both cells is sharper. Under Assumption 4, binary $Z^\star$, and a symmetric, site-invariant channel, the model has five parameters $(q_0, q_1, \varepsilon, \mu_0, \mu_1)$ for the six observed degrees of freedom $P(Z = 1 \mid e)$ and $P(Y = 1 \mid z, e)$; it is therefore over-identified, whereas the classical latent-class model for diagnostic error rates without a gold standard, which it resembles, is exactly identified [33]. For each ε the case mix is fixed by the prevalences and (μ_0, μ_1) must fit four cell rates; in the example the fit is exact only at $\varepsilon = 0.10$ (Section S5.1). Within this parametric model, then, Assumption 4 identifies the error rate $\varepsilon \in [0, \frac{1}{2})$, provided $q_0 \neq q_1$ and $\mu_0 \neq \mu_1$. Without Assumption 4 it does not: M_2 is binary with a symmetric channel ($\varepsilon = 0$) and fits the same table exactly. An external estimate of the error rate, for example from a validation study, is what gives the analysis evidential force: an error rate credibly different from the value implied under Assumption 4 counts against the conjunction of Assumption 4 and the parametric model. The analysis does not remove the identification problem; it relocates it to an interpretable, externally arguable quantity. The calculation is exact for binary $Z^\star$ with a site-invariant, symmetric channel; general sensitivity models are open (main Section 6).

Table S5: Largest cross-site gap in the $Z = 1$ cell that biological invariance can produce, as a function of a posited symmetric representation error rate ε (binary model; observed gap 0.096). The one-cell bound is a necessary condition only; see text.

ε	0.01	0.02	0.05	0.063	0.10	0.15
Maximum gap $\lvert\pi_1(1) - \pi_0(1)\rvert$	0.014	0.029	0.075	0.096	0.160	0.260
Compatible with the gap (one-cell bound)?	No	No	No	Equal	Yes	Yes

S5.1 Exact derivation

All quantities below are exact. With $q_0 = 3/10$, $q_1 = 7/10$, $a_1 = P(Z = 1 \mid Z^\star = 1) = 9/10$, $b_1 = P(Z = 1 \mid Z^\star = 0) = 1/10$, $\mu_0 = 1/5$, and $\mu_1 = 4/5$:

Representation prevalence. $P(Z = 1 \mid E = 0) = \frac{9}{10} \cdot \frac{3}{10} + \frac{1}{10} \cdot \frac{7}{10} = \frac{17}{50}$ and $P(Z = 1 \mid E = 1) = \frac{9}{10} \cdot \frac{7}{10} + \frac{1}{10} \cdot \frac{3}{10} = \frac{33}{50}$.

Posteriors. $\pi_0(1) = \frac{27/100}{17/50} = \frac{27}{34}$, $\pi_1(1) = \frac{63/100}{33/50} = \frac{21}{22}$, $\pi_0(0) = \frac{3/100}{33/50} = \frac{1}{22}$, and $\pi_1(0) = \frac{7/100}{17/50} =$

$\frac{7}{34}$.

Outcome rates. $P(Y=1 \mid Z=1,E=0) = \frac{1}{5} + \frac{3}{5}\cdot\frac{27}{34} = \frac{23}{34}$; $P(Y=1 \mid Z=1,E=1) = \frac{1}{5}+\frac{3}{5}\cdot\frac{21}{22} = \frac{17}{22}$; $P(Y=1 \mid Z=0,E=0) = \frac{1}{5}+\frac{3}{5}\cdot\frac{1}{22} = \frac{5}{22}$; $P(Y=1 \mid Z=0,E=1) = \frac{1}{5}+\frac{3}{5}\cdot\frac{7}{34} = \frac{11}{34}$.

Gaps. $\frac{17}{22} - \frac{23}{34} = \frac{11}{34} - \frac{5}{22} = \frac{18}{187} \approx 0.0963$, equal to $\frac{3}{5}\left(\frac{21}{22} - \frac{27}{34}\right) = \frac{3}{5}\cdot\frac{30}{187}$, as required by Eq. (S8).

Joint distributions. M_1 gives $P(Y=1, Z=1 \mid E=0) = \frac{23}{34}\cdot\frac{17}{50} = \frac{23}{100}$, and similarly $P(Y=0, Z=1 \mid E=0) = \frac{11}{100}$, $P(Y=1, Z=0 \mid E=0) = \frac{3}{20}$, $P(Y=0, Z=0 \mid E=0) = \frac{51}{100}$, and at $E=1$ the values $\frac{51}{100}$, $\frac{3}{20}$, $\frac{11}{100}$, $\frac{23}{100}$. Under M_2, each entry is $P_{M_2}(Y=y \mid Z^\star=z,e)\,P_{M_2}(Z^\star=z \mid e)$ with the values of Step 7, which reproduces the same eight numbers by construction (Lemma 1).

Power calculation. For proportions 23/34 and 17/22, the standard two-sided two-proportion sample-size formula with pooled variance under the null, $\alpha = 0.05$ and power 0.80, gives approximately 337 patients per site in the $Z=1$ stratum.

Sensitivity calculation. For a symmetric error rate ε, $q_e = (P(Z=1 \mid e) - \varepsilon)/(1-2\varepsilon)$ and $\pi_e(1) = (1-\varepsilon)q_e/P(Z=1 \mid e)$. The difference $\pi_1(1) - \pi_0(1)$ equals 0.0144, 0.0291, 0.0753, 0.1604, and 0.2597 at $\varepsilon = 0.01$, 0.02, 0.05, 0.10, and 0.15, and it equals the observed gap 18/187 at $\varepsilon \approx 0.063$. The same values hold in the $Z=0$ cell. Fitting μ_0 and μ_1 to all four cell rates by least squares, with q_e fixed by the prevalences, gives an exact fit only at $\varepsilon = 0.10$ ($\mu_0 = 0.2$, $\mu_1 = 0.8$); the largest absolute residual is 0.031 at $\varepsilon = 0.05$, 0.025 at $\varepsilon = 0.063$, and 0.019 at $\varepsilon = 0.12$.

S6 Extended discussion

S6.1 Clinical representation learning, multisite studies, and external validation

External validation is often read as a test of whether a model has learned something general about patients. Theorem 3 shows that, for representations built on selectively measured data, the result of such a validation cannot by itself say whether that something is biological. A representation that transfers well may do so because sites share mechanisms, or because site-specific mechanisms and site-specific case mix cancel within representation cells (Proposition 1). A representation that transfers poorly may do so because mechanisms differ, or because identical mechanisms are seen through a lossy representation of populations with different case mix (Proposition 2). This does not diminish the value of external validation for its primary purpose, estimating performance in a new setting. It limits what external validation licenses as a statement about mechanism. A representation that survives repeated invariance tests across several independent sites and laboratory types has passed a real and informative test; it has not been shown to isolate biology. The distinction mirrors the philosophy-of-science asymmetry between corroboration and confirmation.

S6.2 Invariant prediction and domain generalization

Methods that exploit heterogeneity across environments either treat invariance as a signature of causal structure [24, 25] or seek predictors robust to bounded shifts [26]. Our results do not contradict those methods on their own terms; they identify a setting in which the premise needs an explicit auxiliary assumption. When the invariance is sought in a learned representation rather than in the latent physiological state, a stable relationship can arise without mechanism invariance, and an unstable one can arise despite mechanism invariance, whenever case mix differs and the

representation is lossy. Invariance-seeking methods applied to clinical embeddings should therefore treat representation sufficiency (Assumption 6b) as a stated premise rather than an implicit one.

S6.3 Algorithmic fairness and health equity

The same construction bears on the equity concerns raised by Obermeyer and colleagues and by Vyas, Eisenstein, and Jones [12, 13]. When sites, or the populations they serve, differ in case mix and in care, Theorem 3 implies that a cross-site difference in how a representation relates to outcomes cannot be attributed from observational data to biology rather than to care, or vice versa. Two failure modes follow. A difference produced by unequal care can be read as a biological difference between populations, which is the pattern by which institutional inequity becomes encoded as apparent biology [13]. Conversely, a genuine difference in the physiology of the populations served can be misread as a practice effect. The representation offers no internal marker that distinguishes these cases. Consequently, claims that attribute group or site differences in learned clinical representations to biology require justification from outside the observational data, such as mechanistic knowledge, interventional evidence, or explicit sensitivity analysis, and better internal validation cannot substitute for that justification.

S6.4 What remains actionable

Three forms of evidence survive Theorem 3 intact, and a fourth lies outside its scope.

Falsification and corroboration. Repeated survival of the invariance test (main Section 4.2) across many independent site pairs and laboratory types corroborates the conjunction of Assumptions 4, 6a, and 6b. A reliable failure falsifies that conjunction without localizing which conjunct failed.

Sensitivity analysis. Equations (S6) to (S8) define a sensitivity model. For posited values of case-mix contrasts and representation error rates, they give the cross-site gap that biological invariance alone would produce; conversely, they yield the smallest representation loss under which case mix could explain an observed gap (Section S5). This converts an unanswerable binary question into a bounded, reportable quantity, in the spirit of Manski's partial-identification program [18, 19]. The exact calculation given here is for binary physiology; general versions are open (main Section 6).

Routine-order anchoring. For protocolized, near-universally ordered measurements, the decision to order carries little information about physiology or about site practice. Anchoring invariance tests on such measurements removes the site-specific ordering pathway identified in Corollary 2. It does not, by itself, remove the case-mix pathway of Proposition 2, which operates even when every patient is measured, so it reduces but does not eliminate reliance on Assumption 6b. With an exact summary of the outcome-relevant state, or two conditionally independent tests at sites with different case mix, it identifies Assumption 4 (Propositions 4 and 5).

Additional identifying variation. Randomized or protocol-driven variation in ordering practice, such as a stepped-wedge change in ordering guidelines or a natural experiment arising from a policy change, supplies information that is not contained in the observational distribution analyzed here. Whether and under what conditions such variation identifies the mechanism is not established in this paper (main Section 5.3), but it is the most direct route across the boundary.

S6.5 Relationship to other impossibility results, and the dark-matter analogy

Formal impossibility results have tended to redirect rather than end the fields that produced them. Bell's theorem is the clearest example of the logical type of our results. Bell derived an observable constraint, an inequality on measurable correlations, that is implied by an entire class of latent-variable models, converting a question about unobservable hidden variables into an empirical one [39]. Our results are statements of the same kind, about which latent-variable models are compatible with which observable distributions, but they land on the opposite side. Biological invariance, unlike Bell's local hidden variables, implies no observable constraint on its own; an observable constraint appears only when it is conjoined with auxiliary conditions on the representation (Theorem 2), and the conjunction can then be falsified but not decomposed. Manski's partial-identification program responded to failures of point identification by building a theory of the bounds that remain [18, 19], and identifiability results in domain adaptation, which hold only under explicit structural restrictions, clarified which additional assumptions any working method must supply [27]. We place the present contribution in the same tradition: a precise negative result that redirects the research program it concerns.

The title's metaphor is meant in a specific sense. In cosmology, dark matter is not observed directly; it is inferred from its gravitational effects on what can be observed, and its properties are constrained by triangulating across independent observational channels rather than by isolating it in any one of them. Clinical judgment stands in a similar relation to the EHR. The latent ordering intent $S^\star$ and the practice component of the outcome mechanism are never recorded, yet they shape every record: which values exist (Theorem 1), which patients fall into which representation cell, and how those cells relate to outcomes (Theorem 3). Our results suggest that clinical judgment should be treated in the same way, as an object to be constrained indirectly by triangulating across sites, measurement types, and sources of interventional variation, rather than as a nuisance to be modeled away within a single observational dataset.

References

[1] Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976. doi: 10.1093/biomet/63.3.581.

[2] Denis Agniel, Isaac S. Kohane, and Griffin M. Weber. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*, 361: k1479, 2018. doi: 10.1136/bmj.k1479.

[3] Rimma Pivovarov, David J. Albers, Jorge L. Sepulveda, and Noémie Elhadad. Identifying and mitigating biases in EHR laboratory tests. *Journal of Biomedical Informatics*, 51:24–34, 2014. doi: 10.1016/j.jbi.2014.03.016.

[4] Rolf H. H. Groenwold. Informative missingness in electronic health record systems: the curse of knowing. *Diagnostic and Prognostic Research*, 4(1):8, 2020. doi: 10.1186/s41512-020-00077-0.

[5] Matthew Phelan, Nrupen A. Bhavsar, and Benjamin A. Goldstein. Illustrating informed presence bias in electronic health records data: how patient interactions with a health system

can impact inference. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*, 5(1):22, 2017. doi: 10.5334/egems.243.

[6] Benjamin A. Goldstein, Nrupen A. Bhavsar, Matthew Phelan, and Michael J. Pencina. Controlling for informed presence bias due to the number of health encounters in an electronic health record. *American Journal of Epidemiology*, 184(11):847–855, 2016. doi: 10.1093/aje/kww112.

[7] Christopher M. Sauer, Li-Ching Chen, Stephanie L. Hyland, Armand Girbes, Paul Elbers, and Leo A. Celi. Leveraging electronic health records for data science: common pitfalls and how to avoid them. *The Lancet Digital Health*, 4(12):e893–e898, 2022. doi: 10.1016/S2589-7500(22)00154-6.

[8] Sarah C. Rossetti, Patricia C. Dykes, Chris Knaplund, Sandy Cho, Jennifer Withall, Graham Lowenthal, et al. Real-time surveillance system for patient deterioration: a pragmatic cluster-randomized controlled trial. *Nature Medicine*, 31(6):1895–1902, 2025. doi: 10.1038/s41591-025-03609-7.

[9] Justin Kauffman, Riccardo Miotto, Eyal Klang, Anthony Costa, Beau Norgeot, Marinka Zitnik, et al. Embedding methods for electronic health record research. *Annual Review of Biomedical Data Science*, 8(1):563–590, 2025. doi: 10.1146/annurev-biodatasci-103123-094729.

[10] Samuel G. Finlayson, Adarsh Subbaswamy, Karandeep Singh, John Bowers, Annabel Kupke, Jonathan Zittrain, et al. The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine*, 385(3):283–286, 2021. doi: 10.1056/NEJMc2104626. Letter.

[11] Adarsh Subbaswamy and Suchi Saria. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics*, 21(2):345–352, 2020. doi: 10.1093/biostatistics/kxz041.

[12] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019. doi: 10.1126/science.aax2342.

[13] Darshali A. Vyas, Leo G. Eisenstein, and David S. Jones. Hidden in plain sight: reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, 383(9):874–882, 2020. doi: 10.1056/NEJMms2004740.

[14] Milena A. Gianfrancesco, Suzanne Tamang, Jinoos Yazdany, and Gabriela Schmajuk. Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, 178(11):1544–1547, 2018. doi: 10.1001/jamainternmed.2018.3763.

[15] Stephen R. Cole and Miguel A. Hernán. Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6):656–664, 2008. doi: 10.1093/aje/kwn164.

[16] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016. doi: 10.1073/pnas.1510507113.

[17] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2nd edition, 2009.

[18] Charles F. Manski. Nonparametric bounds on treatment effects. *American Economic Review*, 80(2):319–323, 1990.

[19] Charles F. Manski. *Partial Identification of Probability Distributions*. Springer, New York, 2003.

[20] Paul L. Williams and Randall D. Beer. Nonnegative decomposition of multivariate information. arXiv:1004.2515, 2010.

[21] Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014. doi: 10.3390/e16042161.

[22] Virgil Griffith and Christof Koch. Quantifying synergistic mutual information. In Mikhail Prokopenko, editor, *Guided Self-Organization: Inception*, volume 9 of *Emergence, Complexity and Computation*, pages 159–190. Springer, Berlin, Heidelberg, 2014. doi: 10.1007/978-3-642-53734-9_6.

[23] Young Sang Choi, Vincent Jeanselme, Pierre Elias, and Shalmali Joshi. ICYM^{2}I: The illusion of multimodal informativeness under missingness. In *International Conference on Learning Representations (ICLR)*, pages 84620–84648, 2026. arXiv:2505.16953.

[24] Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):947–1012, 2016. doi: 10.1111/rssb.12167.

[25] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. arXiv:1907.02893, 2019.

[26] Dominik Rothenhäusler, Nicolai Meinshausen, Peter Bühlmann, and Jonas Peters. Anchor regression: heterogeneous data meet causality. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(2):215–246, 2021. doi: 10.1111/rssb.12398.

[27] Kun Zhang, Bernhard Schölkopf, Krikamol Muandet, and Zhikun Wang. Domain adaptation under target and conditional shift. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 819–827. PMLR, 2013.

[28] Saskia le Cessie, Jan Debeij, Frits R. Rosendaal, Suzanne C. Cannegieter, and Jan P. Vandenbroucke. Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology*, 23(4):551–560, 2012. doi: 10.1097/EDE.0b013e318254f5de.

[29] Tyler J. VanderWeele, Linda Valeri, and Elizabeth L. Ogburn. The role of measurement error and misclassification in mediation analysis. *Epidemiology*, 23(4):561–564, 2012. doi: 10.1097/EDE.0b013e318258f5e4.

[30] Manabu Kuroki and Judea Pearl. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2):423–437, 2014. doi: 10.1093/biomet/ast066.

[31] Wang Miao, Zhi Geng, and Eric J. Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018. doi: 10.1093/biomet/asy038.

[32] Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37(6A): 3099–3132, 2009. doi: 10.1214/09-AOS689.

[33] Siu L. Hui and Stephen D. Walter. Estimating the error rates of diagnostic tests. *Biometrics*, 36(1):167–171, 1980. doi: 10.2307/2530508.

[34] Karthika Mohan, Judea Pearl, and Jin Tian. Graphical models for inference with missing data. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 1277–1285, 2013.

[35] Karthika Mohan and Judea Pearl. Graphical models for processing missing data. *Journal of the American Statistical Association*, 116(534):1023–1037, 2021. doi: 10.1080/01621459.2021.1874961.

[36] Boris S. Mityagin. The zero set of a real analytic function. *Mathematical Notes*, 107(3–4): 529–530, 2020. doi: 10.1134/S0001434620030189.

[37] Christopher Meek. Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 411–418. Morgan Kaufmann, 1995.

[38] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition, 2000.

[39] John S. Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1(3):195–200, 1964. doi: 10.1103/PhysicsPhysiqueFizika.1.195.