Dirty Beakers, Disappearing Government Data, and the Limits of Experimental Inference

Tamar Zeilberger*

December 18, 2025

Abstract

Randomized experiments are widely regarded as the benchmark for causal inference. This paper identifies a primitive inferential failure that arises when nominal assignment does not instantiate a stable causal object. We call this failure the Dirty Beaker Problem. In social science and related fields, the problem takes a particularly intractable form based on how interventions are endogenously transformed within units, so that assignment need not index repeated instances of a unit invariant intervention even under ideal randomization. In such settings, randomization removes selection bias but need not identify the effect of a unit invariant intervention. The paper formalizes this failure and shows that auxiliary information can restrict the set of admissible mappings from assignment to realized intervention without resolving the underlying indeterminacy. Since 2025, the ongoing deletion, alteration, and disruption of U.S. federal datasets have eroded a primary source of such constraint, resulting in experimental contrasts that remain numerically well defined while becoming increasingly indeterminate as causal objects.

Introduction

In 2025, U.S. federal agencies removed, restricted, altered, and delayed access to hundreds of datasets spanning public health and health care, demography, labor markets and income, food security and nutrition, education, environmental and climate monitoring, energy production and use, housing, transportation, and social vulnerability.[1–12] Large portions of the records

*School of Public Policy, London School of Economics and Political Science. Corresponding author: t.zeilberger@lse.ac.uk

that have long structured empirical research are no longer available in consistent or transparent form, with no clear private or ad hoc substitutes.[1]

Researchers across fields have begun to document how these changes disrupt observational research and reduce the set of questions that it can address.[1, 3, 4, 8, 9] This paper considers the wider empirical landscape and identifies a quietly destabilizing consequence. The disappearance and degradation of government data infrastructures alters the inferential conditions under which randomized experiments can lay claim to causality.

What Randomization Guarantees and What It Presumes

Randomized experiments occupy a privileged position in empirical research. That status rests on the properties of random assignment. Randomization is widely understood to secure internal validity by breaking the link between treatment assignment and potential outcomes. To see both what this guarantee delivers and what it leaves unresolved, it is useful to be explicit about the assumptions embedded in standard experimental frameworks.

In the potential outcomes framework, randomization guarantees independence between assignment and potential outcomes.[13] Formally, for each unit i, one posits potential outcomes $Y_i(1)$ and $Y_i(0)$, and randomization implies

$$T_i \perp \{Y_i(1), Y_i(0)\}$$

under the randomization distribution. This result addresses selection bias. It does not, however, guarantee that the potential outcomes themselves correspond to a stable or separable intervention.

The existence and interpretability of potential outcomes require additional assumptions about how interventions operate. In particular, they require that the intervention indexed by assignment can be specified independently of the unit to which it is applied, except through a stable mapping. These assumptions are typically left implicit in discussions of experimental credibility. Importantly, when units are not identical, any differences are assumed to enter as perturbations around a common causal input. That is, assignment must index an intervention whose meaning is invariant across units up to noise.[14]

In laboratory sciences, this distinction is uncontroversial. Chemists do not perform experiments by randomizing over dirty beakers in lieu of cleaning them, because residue interacts with new compounds and becomes part of the reaction itself. Even if residue were perfectly balanced across beakers,

the resulting experiment would remain uninformative about the compound itself.

Outside the laboratory, clean beakers are often not an option. In medicine, beakers are human bodies. They are irreducibly dirty, but biological regularities constrain how treatments interact with underlying physiology. As a result, variation across bodies often affects the magnitude or pathway of a response rather than redefining the treatment itself, and averaging across patients can recover meaningful effects.

In social science and related applications, the beakers are humans. Here, the residue consists of identities, beliefs, histories, norms, incentives, expectations, and so on. These elements form internally interacting systems rather than collections of separable attributes. Assignment enters these systems, so the realized intervention is jointly produced by assignment and internal structure.

The central point here is not that interventions in these settings have heterogeneous effects. It is that nominal assignment may fail to correspond to a repeated causal input in the first place. When the realized intervention is co-produced by the unit into which it is introduced, randomization can eliminate selection bias while leaving the causal object indexed by assignment underdetermined.

The Dirty Beaker Problem: Formalization

We refer to this failure as the *Dirty Beaker Problem*. The Dirty Beaker Problem arises in experimental settings in which the substrate into which an intervention is introduced transforms that intervention in ways that randomization cannot neutralize.

The problem considered here does not arise from interference across units. Units are assumed not to affect one another's outcomes. The difficulty arises because nominal assignment may fail to define a common causal input at the unit level, even in the absence of spillovers.

Let each individual i be characterized by a latent internal state U_i, not observed by the researcher, which includes latent attributes and the internal interaction structure governing how those attributes jointly shape responses to interventions. Let $T_i \in \{0, 1\}$ denote nominal assignment.

Assignment does not act directly on outcomes. Instead, it is transformed into a realized intervention through interaction with the latent state of the

individual:

$$D_i = \phi(T_i, U_i),$$
$$Y_i = g(D_i, U_i).$$

The mapping ϕ captures the endogenous formation of realized interventions through interaction with latent internal states. This is not a claim about heterogeneous effects of a common treatment, but about the absence of unit invariant treatment.

Randomization ensures independence between T_i and latent states. It does not, on its own, ensure that T_i defines stable causal input. The experimental contrast therefore compares collections of realized interventions rather than instances of a unit invariant causal input.

Why Additivity Does Not Rescue Experimental Isolation

At this point it is useful to be explicit about what is being assumed and what is being rejected. Even if latent residue entered outcomes purely additively, a dirty beaker experiment would not isolate the effect of the intervention abstracted from the substrate through which it is realized. It would isolate only the effect of the intervention as realized through the residue present in the substrate. This is why experimental scientists do not randomize over dirty beakers: randomization can, in principle, balance residue, but it cannot remove it, and thus cannot isolate the causal input of interest.

The stronger point made here is that in social systems, and those with analogous features, the residue is not merely additive. Within individuals, assignment encounters accumulated histories of experiences, beliefs, identities, expectations, incentives, and so forth, which interact with one another rather than entering independently. These interactions are not naturally ordered or decomposable, and assignment therefore need not correspond to a stable transformation of a common causal input. The consequence is not simply heterogeneous effects, but the absence of a unit invariant intervention indexed by assignment.

The combinatorial argument below formalizes this distinction and examines what internal interaction structure implies for the repeatability of the realized interventions induced by nominally identical assignments, even under ideal randomization.

Internal Interaction Structure and Combinatorial Indeterminacy

Suppose the latent state of individual i consists of K components,

$$U_i = (U_{i1}, \ldots, U_{iK}),$$

together with an internal interaction structure A_i that governs how these components influence one another. For concreteness, A_i may be represented as a directed adjacency matrix encoding pairwise interactions among components. Even restricting attention to binary interactions with no self edges, the number of distinct internal interaction structures is

$$|\mathcal{A}| = 2^{K(K-1)}.$$

Let the realized intervention be given by

$$D_i = \phi(T_i, U_i, A_i),$$

where ϕ is unrestricted. For fixed K, dependence of ϕ on A_i implies that nominal assignment induces a combinatorially large space of realized interventions.

Because A_i is unobserved and no structure is imposed on how it enters ϕ, a combinatorially large set of distinct mappings from nominal assignment to realized intervention remain compatible with the same experimental design and the same joint distribution of observed outcomes. Nominal assignment therefore need not index any uniquely defined causal input at the unit level.

Crucially, this indeterminacy does not collapse with larger samples or more precise estimation of the joint distribution of (T, Y). The problem is not that the experimental contrast is noisy, but that assignment induces a mixture over realized interventions whose composition is not identified from (T, Y). Additional data can stabilize experimental contrasts without stabilizing the causal input they are taken to represent.

Randomization ensures independence between assignment and latent states, but it does not determine how assignment is transformed into a realized intervention. The inferential problem is therefore not one of estimation error, but of underdetermination of the causal object indexed by assignment.

Estimand Indeterminacy under Dirty Beakers

Although outcomes are generated by realized interventions D_i, experimental estimands are conventionally indexed by nominal assignment T_i rather than by D_i.

In standard experimental analyses, the experimental contrast is interpreted as an estimator of a population level causal estimand,

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)].$$

In dirty beaker systems, this quantity need not correspond to the effect of a unit invariant intervention indexed by nominal assignment, because assignment induces a distribution over realized interventions whose support is large and uncharacterized.

A randomized experiment recovers the sample contrast

$$\hat{\tau}_S = \mathbb{E}_{i \in S}[Y_i \mid T_i = 1] - \mathbb{E}_{i \in S}[Y_i \mid T_i = 0],$$

where S denotes the realized experimental sample. This quantity is well defined numerically. However, numerical definition alone does not imply a well defined causal interpretation, because the mapping from assignment to realized intervention is underdetermined.

One could define the estimand as the effect of assignment to the regime $T = 1$. However, this move does not resolve the problem identified here, because assignment to a regime does not determine which features of the induced distribution of realized interventions are causally operative. The source of the problem lies within units, but its inferential consequence is a failure of across unit comparability.

Auxiliary Information and Causal Restriction

Standard approaches to imperfect correspondence between assignment and treatment presuppose that nominal assignment indexes a unit invariant causal input. Violations of SUTVA, multiple versions of treatment, noncompliance, and estimand redefinition such as LATE all operate by refining inference once the causal input is taken as given.[13–17][1] In dirty beaker systems, this presupposition fails. Nominal assignment need not correspond to a unit invariant intervention. As a result, approaches that redefine treatments, estimands, or target populations do not resolve the problem identified here.

Our central inferential problem is therefore not estimation conditional on assignment, but whether nominal assignment instantiates a determinate

[1] Recent work develops estimand frameworks and aggregation rules for settings with imperfect correspondence between assignment and treatment [18, 19]. These approaches refine inference conditional on a well defined intervention and therefore do not address failures of treatment instantiation.

causal object.[2] In dirty beaker systems, nominal assignment does not act directly on outcomes. Instead, assignment is transformed into a realized intervention through a mapping ϕ that depends on latent internal states and their interaction structure. As a result, nominally identical assignments induce a distribution of realized interventions rather than repeated instances of a unit invariant causal input. Randomization guarantees independence between assignment and latent internal structure, but it does not determine how assignment is realized within units.

Taken on its own, this indeterminacy is severe. When nominal assignment does not index a repeatable causal input, an assignment-indexed experimental contrast does not, by itself, admit a determinate causal interpretation. Balance is achieved, but balance over what remains unspecified. Nothing in the experimental design alone fixes which causal object, if any, an observed contrast represents.

Latent internal states and their interaction structure (U, A) are not observed at the level required to determine how nominal assignment is transformed into realized interventions. Consequently, the joint distribution of (T, Y) is insufficient to fix the causal object indexed by assignment. Information external to the experimental design is therefore required to restrict the set of admissible mappings from assignment to realized intervention.

Absent such auxiliary constraints, the mapping from assignment to realized intervention remains underdetermined even as experimental contrasts stabilize as numerical objects. Increasing sample size does not resolve this indeterminacy. Randomization is orthogonal to the latent interaction structure, so repeated assignment does not collapse the space of realized interventions compatible with a given experimental contrast. As shown in the appendix, without auxiliary restrictions on how assignment is transformed into a realized intervention, the intended treatment effect is not identified even under ideal randomization.

The force of this indeterminacy depends on the extent to which information external to the experimental design restricts the set of admissible mappings from assignment to realized intervention. Where such auxiliary constraints are strong, the range of causal objects compatible with an assignment-indexed contrast may be limited. Where they are weak or

[2]Work on causal identification has examined ill defined interventions, identification under alternative maintained assumptions, and the limits of randomization in a variety of settings [27–29]. These analyses address how causal effects are defined and identified once a causal object is specified. The argument here concerns a logically prior failure. In some settings, nominal assignment may fail to instantiate a unit invariant causal object in the first place, so that identification questions do not yet arise.

absent, the indeterminacy identified above remains unconstrained.

In practice, when auxiliary constraints have entered applied work, it has been through approaches that combine experimental and nonexperimental evidence or emphasize triangulation across methods to address other inferential concerns.[18–26] These approaches are not designed to address failures of treatment instantiation and rather proceed under the maintained assumption that the intervention indexed by assignment is well specified. Nonetheless, by imposing background regularities or constraints not contained in the experimental contrast itself, they can restrict which causal accounts remain admissible.

In this limited sense, auxiliary information can be understood as providing a form of partial cleaning. It does not eliminate the internal residue through which interventions are transformed, but it bounds its influence by ruling out some mappings from assignment to realized intervention. Because these constraints entered for purposes other than treatment instantiation, their effects were partial and uneven, reflecting the incidental manner in which auxiliary restrictions entered practice. When these sources of constraint are misaligned with the causal question of interest, weakened, or disappear, the indeterminacy associated with assignment-indexed contrasts reasserts itself in full.

Replication, Aggregation, and Their Limits

When these auxiliary constraints are unavailable to discipline assignment-indexed contrasts, the question arises whether procedures that do not introduce constraints on how assignment is transformed into realized intervention can do so. Replication and aggregation are natural candidates. The analysis below shows that these procedures inherit the indeterminacy identified above.

Both replication and aggregation presuppose that assignment-indexed contrasts are interpretable as estimates of a common causal quantity, up to sampling variation or structured moderation. Replication is informative only if repeated implementations target the same causal quantity, or quantities that are meaningfully commensurable under a maintained model of variation, so that variation across estimates can be interpreted as sampling variation around a fixed object. Aggregation similarly presupposes that study-level contrasts are interpretable as estimates of the same causal quantity, or as draws from a well defined distribution over such quantities, permitting pooling to refine inference about that object. These are implicit

conditions in the inferential logic of replication and aggregation.

In dirty beaker systems, these presuppositions need not be satisfied. Nominal assignment need not determine how an intervention is instantiated within units. As shown above, the same assignment-indexed contrast is compatible with an underdetermined set of causal quantities, which may include no unit invariant causal quantity at all. In such settings, assignment-indexed contrasts need not correspond to realizations of a single causal quantity. The design and the joint distribution of assignment and outcomes generated by the experiment do not fix what quantity, if any, is being estimated.

Replication does not resolve this indeterminacy. Reimplementing a design yields additional realizations of the joint distribution of assignment and outcomes, but it does not restrict the set of causal quantities compatible with assignment-indexed contrasts. Replications can therefore converge on the same numerical contrast while corresponding to different causal quantities, or to contrasts for which no common causal quantity exists. Agreement across replications establishes stability of assignment-indexed differences, not stability of what those differences are differences of. Disagreement across replications is equally uninformative, because it does not distinguish sampling variation from variation in the causal quantity being instantiated.

Aggregation, including meta-analytic pooling and related averaging procedures, does not resolve this indeterminacy. Each study-level contrast is compatible with an underdetermined set of causal quantities, which may differ across studies and need not include any shared causal quantity. Pooling presupposes that the contrasts being combined admit a common estimand, or are draws from a well defined distribution over such estimands. In dirty beaker systems, that presupposition need not hold. Aggregation therefore applies an averaging operation to contrasts whose associated causal quantities need not be jointly defined. The resulting aggregate cannot be interpreted as an estimator of any determinate causal quantity under the information supplied by the experimental designs.

Importantly, aggregation does not merely fail to recover a common estimand. By mapping a collection of underdetermined contrasts into a single numerically stable summary, it produces an object whose relationship to admissible causal quantities is weaker than that of the individual contrasts from which it is constructed. In this sense, aggregation can worsen the indeterminacy of causal interpretation, even as numerical stability increases.

The difficulty is not heterogeneity around a shared effect. It is the absence of a shared estimand to which replication or aggregation could coherently apply. When assignment does not instantiate a determinate causal quantity, procedures that presuppose such a quantity cannot supply it.

Replication stabilizes contrasts. Aggregation stabilizes averages of contrasts. Neither stabilizes the causal object those contrasts are taken to represent.

A Corollary for Procedural Guarantees

The results above imply a general constraint on what procedural commitments in randomized experimental research can establish. Procedures that fix assignment mechanisms, estimands, or analysis plans in advance presuppose that nominal assignment instantiates a determinate causal object. They do not, on their own, determine how assignment is transformed into a realized intervention within units.

When the mapping from assignment to realized intervention is underdetermined, committing ex ante to a particular estimand or analysis plan does not fix the causal object that the resulting contrast represents. Such commitments can ensure internal coherence between design, estimation, and reporting, but they cannot guarantee that the contrast produced corresponds to a unit invariant causal input.

This limitation is not a consequence of imperfect implementation or researcher discretion. It follows from the structure of the problem. Even under ideal adherence to procedural guarantees, assignment-indexed contrasts may remain compatible with multiple, many, or no, determinate causal objects. In dirty beaker systems, procedural commitments restrict how contrasts are produced and reported without restricting the set of causal objects consistent with those contrasts.

Data Disappearance and Experimental Inference

Randomized experiments face a general inferential failure whenever nominal assignment does not, on its own, define a stable causal object. This failure is structural, not procedural. In 2025, the empirical conditions that limited, to some unknown extent, the practical consequences of this failure changed materially. U.S. federal agencies began to remove, restrict, alter, and delay access to hundreds of datasets that had long structured empirical research.[1–12]

These records have no private or ad hoc replacement.[1] They were produced at scale, maintained continuously, and applied across populations rather than research samples. Their coverage was mandated rather than voluntary, their continuity institutional rather than project based, and their structure stabilized by administrative practice rather than analytic choice.

For experimental inference, this matters because such records supply auxiliary regularities that no individual experiment can generate and no collection of bespoke studies can reconstruct.

The relevance of this material change does not depend on which auxiliary constraints matter for any particular experiment. In dirty beaker systems, the need for external constraints is structural, while the form those constraints take is experiment specific and unknowable ex ante. What matters is not the loss of any particular dataset, but the contraction of the class of records capable of supplying structure at scale. As these records are removed, restricted, rendered discontinuous, and otherwise compromised, fewer external constraints remain available to restrict how nominal assignment is instantiated within units.

The consequence is mechanical. For a fixed experimental design, the set of admissible mappings from assignment to realized intervention expands. Assignment-indexed contrasts remain numerically well defined and may remain stable under replication and aggregation, but those contrasts are consistent with an expanding set of causal accounts, including accounts under which no unit invariant causal input exists. The erosion of government data infrastructures does not produce a new inferential failure. It renders a primitive inferential failure no longer something we can ignore.

Conclusion

Random assignment removes selection bias. It does not, by itself, ensure that nominal assignment defines a stable causal object. This paper identifies a primitive inferential failure that arises whenever an intervention is transformed by the substrate into which it is introduced. In such settings, nominal assignment may fail to index a unit invariant causal input even under ideal experimental conditions. The problem concerns the existence and interpretability of the estimand itself, and it arises prior to questions of estimation, precision, or replication.

The chemist's refusal to randomize over dirty beakers captures the core logic. Randomization can balance contamination, but it cannot isolate a causal input whose form depends on interactions within the substrate. In social science and related fields, interventions are typically realized through latent, internally interacting systems. Nominally identical assignments therefore need not correspond to repeated instances of a common causal input.

Nothing internal to experimental design can resolve this failure. Asymptotics, replication, and aggregation stabilize assignment-indexed contrasts,

but they do not determine what causal object, if any, those contrasts represent. Information external to the experiment can constrain the set of admissible causal accounts, but it cannot supply what randomization alone does not define.

The erosion of government data infrastructures does not produce this failure. It removes sources of constraint that had limited its practical consequences. Under those conditions, the inferential limits of randomized experiments become harder to ignore. Where nominal assignment does not define a stable causal object, randomized experiments do not, on their own, allow determinate causal claims.

Appendix: Non-identification of the Intended Treatment Effect in Dirty Beaker Systems

This appendix formalizes a claim made in the main text: in dirty beaker systems, random assignment does not identify the causal effect of an intended intervention without additional restrictions on how assignment maps into realized interventions.

A. Identification

Let $\mathcal{M}$ denote a class of data-generating processes indexed by m, each inducing a joint distribution $P_m(T, Y)$ over observable assignment T and outcome Y. Let $\tau^\star(m)$ denote a causal quantity of interest. The class $\mathcal{M}$ places no restrictions on the mapping ϕ beyond measurability.

Definition (Identification). The causal quantity $\tau^\star$ is identified on $\mathcal{M}$ if for any $m_1, m_2 \in \mathcal{M}$,

$$P_{m_1}(T, Y) = P_{m_2}(T, Y) \quad \Rightarrow \quad \tau^\star(m_1) = \tau^\star(m_2).$$

B. Setup

Let $T \in \{0, 1\}$ denote random assignment, independent of a latent state U. Assignment induces a realized intervention $D = \phi(T, U)$, and outcomes satisfy

$$Y = g(D, U).$$

Let d_0 denote a baseline intervention and $d^\star$ an intended treated intervention. Define the intended treatment effect as

$$\tau^\star(m) = \mathbb{E}_m[g(d^\star, U) - g(d_0, U)].$$

C. Proposition

Proposition. The intended treatment effect $\tau^\star$ is not identified from the joint distribution of (T, Y) on the class $\mathcal{M}$ of dirty beaker systems in which assignment induces realized interventions through an unobserved latent state (or more generally, a latent state not fully characterized by observables).

D. Proof

Fix any $p \in (0, 1)$. Let $U \sim \text{Bernoulli}(p)$ and $T \sim \text{Bernoulli}(1/2)$, independent of U. Let $Y \in \{0, 1\}$.

Construct two models $m_A = (\phi_A, g_A)$ and $m_B = (\phi_B, g_B)$ in $\mathcal{M}$.

Model A.

$$\phi_A(0, U) = d_0, \qquad \phi_A(1, U) = d^\star,$$

$$g_A(d_0, U) = 0, \qquad g_A(d^\star, U) = U.$$

Model B. Let $d' \neq d^\star$ and define

$$\phi_B(0, U) = d_0, \qquad \phi_B(1, 0) = d^\star, \qquad \phi_B(1, 1) = d',$$

$$g_B(d_0, U) = 0, \qquad g_B(d^\star, U) = 0, \qquad g_B(d', U) = \mathbf{1}\{U = 1\}.$$

In both models,

$$\mathbb{E}[Y \mid T = 0] = 0, \qquad \mathbb{E}[Y \mid T = 1] = p,$$

so $P_{m_A}(T, Y) = P_{m_B}(T, Y)$. However,

$$\tau^\star(m_A) = p, \qquad \tau^\star(m_B) = 0.$$

Since two observationally equivalent models imply different values of $\tau^\star$, the intended treatment effect is not identified from randomized data in this class of systems. This formal result reflects the core claim of the paper: randomization removes selection bias but does not, by itself, define a stable causal input. □

E. Auxiliary Constraints and Partial Cleaning

The non-identification result above shows that randomized assignment does not, on its own, identify the intended treatment effect in dirty beaker systems. This subsection clarifies what auxiliary information can and cannot accomplish in such settings.

Let $\mathcal{M}$ denote the class of data-generating processes defined above, and let $P_{Y,T}$ denote the joint distribution of (Y, T) induced by random assignment. Let $\mathcal{M}(P_{Y,T}) \subseteq \mathcal{M}$ denote the set of models observationally equivalent with respect to (T, Y).

Let $\mathcal{H}$ denote auxiliary information external to the experiment that restricts admissible models through constraints on the latent state U, its distribution, or the mapping ϕ from assignment to realized intervention. Define the restricted model class

$$\mathcal{M}(P_{Y,T}, \mathcal{H}) = \mathcal{M}(P_{Y,T}) \cap \mathcal{M}(\mathcal{H}).$$

Then $\mathcal{M}(P_{Y,T}, \mathcal{H}) \subseteq \mathcal{M}(P_{Y,T})$. For any causal query Q defined on $\mathcal{M}$, the corresponding identified set satisfies

$$\mathcal{Q}(P_{Y,T}, \mathcal{H}) \subseteq \mathcal{Q}(P_{Y,T}),$$

where $\mathcal{Q}(P_{Y,T}) = \{Q(m) : m \in \mathcal{M}(P_{Y,T})\}$ and $\mathcal{Q}(P_{Y,T}, \mathcal{H}) = \{Q(m) : m \in \mathcal{M}(P_{Y,T}, \mathcal{H})\}$.

Auxiliary information can therefore restrict the range of causal interpretations compatible with a given experimental contrast without identifying a unique causal object. This set-restriction result corresponds to the partial cleaning discussed in the main text.

F. A Concrete Query: Induced Distributions of Realized Interventions

To make the set-restriction argument concrete, consider the query

$$Q \equiv \left(P(D \mid T = 1),\ P(D \mid T = 0)\right),$$

the pair of induced distributions of realized interventions under treatment and control assignment.

In clean-beaker settings, nominal assignment indexes a unit invariant intervention, and Q collapses to a degenerate or low-dimensional object. In dirty beaker systems, assignment induces a mixture over realized interventions whose composition depends on the latent state U and the mapping ϕ.

Absent auxiliary constraints on U or ϕ, the joint distribution of (T, Y) does not identify Q. Multiple distinct distributions over realized interventions can generate the same experimental contrast. Auxiliary information can restrict the admissible distributions of D under assignment, thereby narrowing the identified set for Q without collapsing it to a singleton.

References

[1] Freilich J, Price WN II, Kesselheim AS (2025) Disappearing Data at the U.S. Federal Government. *New England Journal of Medicine* 392(24). doi:10.1056/NEJMp2502567.

[2] Cox C, Rae M, Kates J, Wager E, Ortaliza J, Dawson L (2025) A Look at Federal Health Data Taken Offline. KFF (Kaiser Family Foundation), February 2, 2025.

[3] Council of Professional Associations on Federal Statistics (2025) Protecting Public Access to Federal Data. February 11, 2025.

[4] Center on Budget and Policy Priorities (2025) Federal Data Are Disappearing, Including Our Best Measure of Household Food Security. September 29, 2025.

[5] Satter R (2025) Harvard law library acts to preserve government data amid sweeping purges. *Reuters*, February 6, 2025.

[6] Reuters (2025) U.S. government shutdown: How it affects key economic data publishing. October 6, 2025.

[7] Stein R (2025) The public lost access to Census Bureau data for days after a Trump order. NPR, February 12, 2025.

[8] STAT News (2025) Interrupted access to unique maternal health database rattles researchers.

[9] Nieman Lab (2025) How do you report on the weather when data is disappearing?

[10] National Security Archive, George Washington University (2025) Disappearing data, Part II: Distorted.

[11] Bureau of Economic Analysis (2025) Discontinued or Delayed Statistics: 2024–2025. U.S. Department of Commerce.

[12] Bureau of Labor Statistics (2025) 2025 federal government shutdown impact on the Consumer Price Index. U.S. Department of Labor.

[13] Imbens GW, Rubin DB (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

[14] Rubin DB (1980) Randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association* 75(371):591–593.

[15] Frangakis CE, Rubin DB (2002) Principal stratification in causal inference. *Biometrics* 58(1):21–29.

[16] Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. *Econometrica* 62(2):467–475.

[17] Angrist JD, Imbens GW, Rubin DB (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434):444–455.

[18] Colnet B, et al. (2024) Causal inference methods for combining randomized trials and observational studies. *Statistical Science* 39(1):165–191.

[19] Rosenman ETR (2025) Methods for Combining Observational and Experimental Causal Estimates: A Review. *Wiley Interdisciplinary Reviews: Computational Statistics* 17(2):e70027. doi:10.1002/wics.70027.

[20] Bareinboim E, Pearl J (2016) Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* 113(27):7345–7352.

[21] Pearl J, Bareinboim E (2014) External validity: From do-calculus to transportability across populations. *Statistical Science* 29(4):579–595.

[22] Cole SR, Stuart EA (2010) Generalizing evidence from randomized clinical trials to target populations. *American Journal of Epidemiology* 172(1):107–115.

[23] Dahabreh IJ, Robertson SE, Stuart EA, Hernán MA (2019) Generalizing causal inferences from randomized trials. *Biometrics* 75(2):685–694.

[24] Deaton A, Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* 210:2–21.

[25] Lawlor DA, Tilling K, Davey Smith G (2016) Triangulation in aetiological epidemiology. *International Journal of Epidemiology* 45(6):1866–1886.

[26] Munafò MR, Davey Smith G (2018) Robust research needs many lines of evidence. *Nature* 553(7689):399–401.

[27] Pearl J (2009) *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press.

[28] Manski CF (2007) *Identification for Prediction and Decision*. Harvard University Press.

[29] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) *Bayesian Data Analysis*. 3rd ed. Chapman and Hall/CRC.