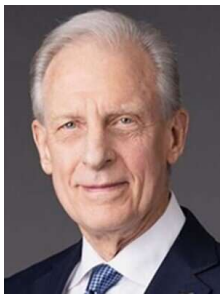
Furtive Changes To Federal Health Data Threaten Admissibility

By **William Lee, Ian Barlow and William Kershaw** (August 14, 2025)

Editor's note: Law360 welcomes opinionated commentary and debate in our Expert Analysis section. To submit op-eds or rebuttals, or to speak to an editor about submissions, please email expertanalysis@law360.com.

William Lee

Ian Barlow



William Kershaw

A recent study published in The Lancet highlights a troubling trend: Nearly 100 U.S. federal health datasets were quietly modified in early 2025 without adequate transparency.[1]

These changes occurred without any notation in the official change logs, undermining bipartisan congressional support for more transparent data policy under the Open, Public, Electronic and Necessary Government Data Act, passed in 2019, and raising red flags about the integrity and admissibility of federal health data in litigation.[2]

The study, authored by Janet Freilich of Boston University School of Law and Aaron Kesselheim of Brigham and Women's Hospital and Harvard Medical School, reveals a systemic pattern of undocumented alterations to U.S. federal health datasets that could carry significant evidentiary consequences.[3]

The authors reviewed 232 datasets and related metadata that were modified by federal agencies between Jan. 20 and March 25 of this year.[4]

These datasets span U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, and U.S. Department of Veterans Affairs data tracking a large number of environmental exposures, confounders and risk factors, health outcomes, and morbidity and mortality.[5]

Of the 232 datasets reviewed, the authors identified 114 datasets, or 49%, with substantial changes — of which 89 datasets, or 78%, were subject to reclassification or recategorization of the data, such as changes to column headers, stratification categories or descriptive elements of the data.[6]

Only 15 datasets, or 13%, logged or otherwise indicated that these changes had occurred, and many lacked any form of transparent version control.[7] The most common change across altered datasets related to gender and sex classifications, which can be associated not only with certain exposures, but also with certain outcomes.[8]

The foregoing should concern plaintiffs counsel, defense counsel and jurists. Litigants on both sides of the aisle routinely seek to admit federal health data as part of the fact record and expert witness testimony, and the courts must gatekeep unreliable evidence from reaching the fact-finder, raising serious questions about the propriety and admissibility of data with undocumented alterations.

Under Federal Rule of Evidence 803(8), public records may be admitted as an exception to the hearsay rule based on an underlying presumption of trustworthiness.[9]

This presumption is predicated on the principle that such records are generated under transparent, neutral and methodical procedures — conditions that inherently reduce the risk of fabrication or error — unless the courts find circumstances indicating a lack of trustworthiness of the information or conclusions drawn therefrom.[10]

Data transparency is vital to its integrity. When the government engages in haphazard manipulation, recharacterization or reclassification of official datasets without recording changes in the metadata, history logs or elsewhere, the data's integrity, authenticity and chain of custody are called into question — potentially jeopardizing the admissibility of such data when proffered as evidence.

Moreover, the reliability and content validity of the data itself may be compromised, as often these post hoc modifications have not been contemplated by the original data collection methodologies or the reviewers coding the raw data.[11]

A classification category added after the fact, or redefined without methodological justification, can transform how an expert interprets population-level findings, incidence rates or exposure-outcome relationships.

Such misclassifications of data can lead to false positives, false negatives, and spurious or speculative results ripe for exclusion. For example, nondifferential misclassification — where exposure or outcome is measured with error equally across all study groups — generally biases risk estimates toward the null.[12]

In a study that measures mortality associated with a certain exposure in women, random gender or sex reclassifications can have the deleterious effect of underestimating the true mortality associated with the exposure of interest — a spurious result with significant litigation and public health consequences.

This lack of transparency is not a matter of semantics. In litigation — particularly cases involving product liability and public health, for which epidemiological analysis is generally the mainstay of establishing causation — experts for both sides often rely on federal health data to support their conclusions.

If the structural underpinnings of that data have shifted without documentation, opposing counsel has a compelling basis to challenge its reliability under the Daubert, Frye or other expert admissibility standards.

Additionally, the courts may decline to apply the public records exception to the underlying data altogether, finding that the circumstances surrounding the record's explanatory structure and chain of custody render it untrustworthy.

Moving forward, federal agencies must strive for transparent and accountable data management, documenting any changes to their datasets in compliance with best scientific practices and applicable laws. Meanwhile, litigants should consider:

- Archiving current and older versions of government datasets contemporaneously with expert reliance;

- Requesting documentation or logs of dataset changes under Freedom of Information Act or discovery mechanisms;

- Ensuring experts scrutinize classification structures for inconsistencies with published methodology; and

- Strategizing around preemptive admissibility challenges that may arise on pretrial motions in limine.

In short, the silent reclassification of public health data by federal agencies — absent proper documentation — goes to the very heart of the hearsay exception under Federal Rule of Evidence 803.

Courts rely on the integrity of public records — and when that integrity is in question, the admissibility of those records and related expert testimony may not survive judicial scrutiny. Effective advocacy for both plaintiffs and defendants depends on transparent data policy.

William J. Lee is leader of the scientific practice group at Kershaw Talley Barlow PC.

Ian J. Barlow is a partner at the firm.

William A. Kershaw is a managing partner at the firm.

[1] See Freilich, J., Kesselheim, A.S., Data manipulation within the US Federal Government, Lancet, July 19, 2025; 406(10500):227-228, doi: 10.1016/S0140-6736(25)01249-8.

[2] See Open, Public, Electronic, and Necessary (OPEN) Government Data Act, Pub. L. No. 115-435, tit. II, 132 Stat. 5529 (2019).

[3] See Freilich and Kesselheim at 228.

[4] Id. at 227.

[5] Id. Confounders and risk factors are inherent characteristics, e.g., genetics, or environmental exposures that can obscure a true association between an exposure of interest and outcome of interest.

[6] Id.

[7] Id.

[8] Id.

[9] See Fed. R. Evid. 803(8).

[10] See Beech Aircraft Corp. v. Rainey, 488 U.S. 153 (1988).

[11] Content validity refers to the extent to which an assessment accurately reflects what it is intended to measure based on attributes such as representativeness, relevance and comprehensiveness.

[12] See, e.g., Wacholder., S, Hartge, P., Lubin, J.H., Dosemeci, M., Non-differential misclassification and bias towards the null: a clarification, Occup Environ Med., 1995 Aug; 52(8):557-8. doi: 10.1136/oem.52.8.557.