

Stress-testing post-market medical-device reporting signals: an empirical study of temporal, entity-label and comparator robustness in FDA MAUDE

Donglin Xu

Qilu Hospital of Shandong University

Zhen Yu

Qilu Hospital of Shandong University

Ang Li

Qilu Hospital of Shandong University

Zengkun Liu

Qilu Hospital of Shandong University

Qiduo Chen

Qilu Hospital of Shandong University

Jing Liu

Qilu Hospital of Shandong University

Hourong Sun

sunhourong2016@126.com

Qilu Hospital of Shandong University

Research Article

Keywords: FDA MAUDE, post-market surveillance, medical device reporting, methodological robustness, disproportionality analysis, comparator sensitivity

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10862018/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Broad-period positive disproportionality does not necessarily indicate analytical robustness. The same post-market reporting signal may change classification when evaluated with shorter time windows, submitted entity-label removal, alternative comparator specifications, or cumulative versus rolling strategies.

Methods

We conducted a retrospective empirical methodological study using FDA MAUDE records received from 2020 through 2025. The primary unit was the unique MDR_REPORT_KEY. The final analysis universe contained 61 product-code/mechanism pairs, of which 19 met formal criteria in both broad periods. We assessed quarterly and leave-one-year-out persistence, manufacturer- and model-label removal, leave-one-comparator-code-out perturbation, 24-cutoff cumulative replay, 21 rolling four-quarter windows, and algorithmic recall linkage for regulatory context.

Results

Five of 19 signals met all temporal and entity-label robustness criteria. Under comparator perturbation, 10 signals remained robust across all evaluated scenarios, four retained direction but lost precision in at least one scenario, and five reversed direction. The five most robust signals had rolling LCL > 1 in 21/21 windows, whereas the predefined fragile group had median persistence 0.857 (rank-biserial = 0.846; two-sided $P = 0.006$; exploratory).

Conclusions

Broad-period consistency alone is insufficient evidence of methodological stability. Temporal, entity-label and comparator stress tests provide complementary sensitivity information, and rolling local analysis can reveal instability masked by cumulative replay. In MAUDE, these findings describe reporting disproportionality rather than incidence, comparative safety, causality, prediction, or independent validation.

Background

The U.S. Food and Drug Administration (FDA) Manufacturer and User Facility Device Experience (MAUDE) database is a public source for post-market spontaneous reporting and surveillance of medical devices [1, 2]. It provides a practical methodological test bed in which reporting patterns can be examined across heterogeneous device technologies and failure mechanisms [3–11].

A reporting odds ratio (ROR) estimated over a broad time period can remain elevated despite quarter-to-quarter fluctuation, concentration within one submitted manufacturer or model label, or dependence on a small part of a comparator definition. Cumulative analyses can also retain historical evidence after local support has weakened. These choices concern analytical specification rather than new biological evidence.

MAUDE has no device-use exposure denominator, and reports may be affected by under-reporting, stimulated reporting, incomplete information and manufacturer reporting practice [1, 2, 12–14]. ROR therefore measures reporting disproportionality rather than incidence, absolute risk, comparative safety or causality. Many reporting-signal analyses emphasize a single or broad time interval, but fewer empirically quantify how perturbing time scale, submitted entity labels, comparator specification and cumulative versus rolling strategy changes the resulting classification.

We therefore evaluated how analytical choices affect the interpretation of a fixed post-market signal universe. The study was designed as an empirical methodological demonstration, not as a validated framework, prediction tool or causal safety assessment. DTZ flow/occlusion, DWF structural fracture, LOX structural fracture and MKJ display/interface were retained as illustrative archetypes within the complete signal landscape rather than as a separate case series.

Methods

Study design and data sources

This retrospective empirical methodological study used publicly available FDA post-market surveillance records. MAUDE records with FDA date-received values from 1 January 2020 through 31 December 2025 were analyzed. Public FDA Medical Device Recalls records were used only for regulatory context [15]. Device, master-event, device-problem and patient-problem records were linked using MDR_REPORT_KEY. The reproducibility archive begins from the compact derived inputs used for the submitted analyses; raw FDA files are not redistributed (Supplementary Figure S1; see Additional file 1).

The primary analytical unit was the unique MDR_REPORT_KEY. Device-problem records were deduplicated by MDR_REPORT_KEY and FDA problem code before phenotype sets were formed. Device rows were deduplicated using the identifiers and normalized submitted labels defined in the analysis. Distinct MDR keys may represent follow-up or supplemental submissions, and reports are not unique patients or independently verified clinical events.

Signal definitions and comparators

The final analysis universe comprised 61 product-code/mechanism pairs. All integrations used product_code plus mechanism as the compound key. Exact phenotypes were sets of official FDA device-problem codes, and signal reports were MDR_REPORT_KEY unions over those exact codes within target or

comparator report sets. Each target product code used its final fixed comparator-code set, held constant across the two broad periods.

Broad-period analysis

FDA date-received years 2020–2023 defined the earlier period and 2024–2025 the later period. For each pair, a was the number of target reports containing the exact phenotype, b target reports without it, and c and d the corresponding comparator counts. $ROR = (a/b)/(c/d)$. When any cell was zero, 0.5 was added to all four cells for the ROR and Wald 95% confidence interval; the one-sided exact Fisher test used the original integer table. Benjamini–Hochberg correction was applied separately to the 61-pair family in each broad period.

The earlier criterion required $a \geq 30$, $ROR > 1$, lower 95% confidence limit (LCL) > 1 and BH-FDR < 0.05 . The later criterion required $a \geq 20$ with the same ROR, LCL and FDR thresholds. Nineteen pairs met both criteria. The selected 61-pair family was defined after upstream screening and refinement; its FDR does not retrospectively cover those upstream decisions.

Temporal and entity-label sensitivity

Quarterly RORs were calculated for 2024Q1–2025Q4. Annual leave-one-year-out (LOO) analyses excluded each year from 2020 through 2025. Entity-removal analyses separately removed the leading exact normalized manufacturer-name label and the leading named normalized model label selected from pooled target-signal reports. These labels are not corporate families or model families.

A deterministic hierarchy produced one class per pair. The quarter temporal gate required LCL > 1 in at least 6 of 8 quarters, and the annual LOO gate required LCL > 1 in all 6 leave-one-year-out analyses. Tier D denoted failure to meet both broad-period criteria. Tier A required formal criteria in both broad periods, the quarter gate, the annual LOO gate, manufacturer-removal results with $ROR > 1$ and LCL > 1 in both periods, and model-removal results with $ROR > 1$ and LCL > 1 in both periods. Tier B required a two-period signal and both temporal gates, but at least one evaluable manufacturer- or model-removal period failed the $ROR > 1$ and LCL > 1 requirement. Tier C required a two-period signal with a failed quarter gate or annual LOO gate. Tier U denoted an otherwise qualifying two-period signal with incomplete required robustness coverage and no observed failure. Missing results were recorded as NA, not as failures. These thresholds are transparent operational stress-test definitions, not externally validated clinical cut points.

Comparator perturbation

For each product code in the final comparator set, leave-one-comparator-code-out (LOCO) analysis removed that code and recomputed the remaining report union and 2×2 table in each broad period. An unavailable post-removal denominator was recorded as NA. Comparator-A required all evaluable LOCO scenarios to have $ROR > 1$ and LCL > 1 , and all genuinely distinct alternative comparator specifications that

existed before this analysis to retain $ROR > 1$ and $LCL > 1$ in both broad periods. Comparator-B required all evaluable scenarios to retain direction with $ROR > 1$ but at least one $LCL \leq 1$. Comparator-C required at least one evaluable scenario with $ROR \leq 1$. Comparator-U denoted incomplete required coverage without observed B/C fragility. No alternative comparator was created for this analysis, and LOCO did not create a new FDR discovery family. Comparator concentration was summarized with fractional top-one shares and HHI but was not a failure criterion (Supplementary Figure S2; see Additional file 1).

Historical replay

Cumulative 2x2 tables were recalculated at 24 cutoffs from 2020Q1 through 2025Q4 using the same definitions, final analysis universe and final fixed comparators. At each cutoff, STATISTICAL_SIGNAL_LOW_COUNT required $ROR > 1$, $LCL > 1$ and cutoff-specific BH-FDR < 0.05 with exact signal count < 20 . EVIDENCE_SUFFICIENT_SIGNAL required exact signal count ≥ 20 , $ROR > 1$, $LCL > 1$ and cutoff-specific BH-FDR < 0.05 . CONFIRMED_PERSISTENT_SIGNAL required the current and immediately preceding cutoffs both to be EVIDENCE_SUFFICIENT_SIGNAL. NO_SIGNAL denoted failure to meet the statistical crossing conditions. Structurally non-estimable results remained NA and were not recoded as NO_SIGNAL. The 61-pair multiplicity family remained fixed at every cutoff.

Local persistence was assessed in 21 rolling four-quarter windows ending 2020Q4 through 2025Q4. These windows were descriptive and did not create sequential alpha-spending or alerts. The replay was conditional on definitions reviewed over the complete 2020–2025 record and was not prospective validation.

Recall linkage

Recall events were deduplicated deterministically using event number, then recall number when event number was absent. The earliest valid initiation date within an event group was primary, with posting date secondary. R1 indicated at least one unique high-confidence event with exact mechanism concordance; R2 indicated product-code recall context without high exact-mechanism concordance. Recall linkage was used as regulatory context, not as a gold standard or prediction test (Supplementary Figure S3; see Additional file 1).

Statistical analysis and reproducibility

The primary analyses used RORs, Wald 95% confidence intervals, one-sided exact Fisher tests and BH-FDR. The most robust group was defined as temporal/entity Tier A plus Comparator-A ($n = 5$). The predefined fragile group was defined after pair-level deduplication as any two-period signal with Tier B or Tier C or Comparator-B or Comparator-C ($n = 13$); the remaining one two-period signal belonged to neither group. The rolling-LCL comparison between these five most robust signals and 13 predefined fragile signals used a two-sided Mann–Whitney test and rank-biserial effect estimate. Because the group definition partly incorporated quarter-level temporal persistence, this comparison was descriptive and exploratory, post-

selection, based on small groups and affected by pair-level non-independence; it was not independent validation of the robustness classification. Recall-group comparison used a two-sided Fisher test and was also exploratory, post-selection and pair-level non-independent. Numbered provenance scripts, the compact evidence workbook and a portable generator are provided in Additional file 2 and Additional file 3.

Use of large language models

OpenAI ChatGPT (GPT-5.6 Sol, OpenAI) was used for language translation and editing and for coding assistance during manuscript preparation. All AI-assisted text and code were reviewed and verified by the authors, who take full responsibility for the final manuscript and analyses.

Results

Broad-period consistency and signal robustness

The final universe contained 61 evaluable product-code/mechanism pairs. Twenty-six met the earlier-period formal criterion and 20 met the later-period criterion; 19 met both. These two-period signals arose from 13 target product codes and are shown within the complete 61-pair landscape in Fig. 1B. Five were Tier A, four Tier B, eight Tier C and two Tier U (Fig. 1C,D). Broad-period consistency therefore defined a focused signal set but did not establish uniform robustness.

The integrated profiles of the 19 two-period signals are summarized in Table 1.

Table 1
Integrated profile of the 19 two-period signals

Code	Mechanism	ROR, earlier / later	Robustness tier	Comparator class	Rolling 4Q persistence	Recall context
DWF ^a	Structural fracture	6.43 / 4.48	Tier A	Comparator-A	21/21	R1
DTZ ^a	Flow/occlusion	3.95 / 8.69	Tier A	Comparator-A	21/21	R1
MHX	Display/interface	3.75 / 6.39	Tier A	Comparator-A	21/21	R1
DQK	Communication	8.64 / 3.26	Tier A	Comparator-A	21/21	R2
LOX ^a	Structural fracture	1.60 / 1.46	Tier A	Comparator-A	21/21	R1
DRS	Flow/occlusion	2.55 / 10.65	Tier B	Comparator-C	11/21	R1
DSI	Display/interface	1.30 / 3.42	Tier B	Comparator-C	17/21	R1
DWF	Other/unclassified	1.88 / 2.07	Tier B	Comparator-A	15/21	R2
DTQ	Other/unclassified	1.44 / 1.87	Tier B	Comparator-A	19/21	R2
MHX	Software/firmware	6.89 / 3.43	Tier C	Comparator-A	20/21	R2
MHX	Communication	3.53 / 1.62	Tier C	Comparator-B	20/21	R2
DQO	Displacement/detachment	16.79 / 39.12	Tier C	Comparator-A	21/21	R2
DQA	Display/interface	2.77 / 1.58	Tier C	Comparator-B	13/21	R1
DSI	Sensing/detection	2.39 / 1.53	Tier C	Comparator-C	18/21	R2
DTZ	Structural fracture	3.09 / 1.47	Tier C	Comparator-B	18/21	R1
MKJ ^a	Display/interface	5.75 / 1.86	Tier C	Comparator-C	6/21	R1
DQA	Sensing/detection	52.49 /	Tier C	Comparator-B	15/21	R2

Code	Mechanism	ROR, earlier / later	Robustness tier	Comparator class	Rolling 4Q persistence	Recall context
		25.19				
DSY	Packaging/sterility	19.56 / 13.78	Tier U	Comparator-A	21/21	R2
DWJ	Other/unclassified	2.85 / 3.27	Tier U	Comparator-C	21/21	R2

^a Indicates the four illustrative archetypes. Rows follow the fixed display order; no weighted score or rank is shown. ROR values are reported as earlier / later period estimates for 2020–2023 and 2024–2025. Rolling 4Q persistence is the number of 21 windows with LCL > 1. R1 denotes high mechanism-concordance recall context and R2 product-code context only. CI, confidence interval; LCL, lower confidence limit; ROR, reporting odds ratio.

Comparator sensitivity

Comparator perturbation separated the 19 signals: 10 were Comparator-A, four Comparator-B and five Comparator-C (Fig. 2A). All five Tier-A signals remained Comparator-A. Nine signals had at least one LOCO scenario with LCL ≤ 1, and five had at least one direction reversal with ROR ≤ 1. Comparator-B therefore reflected loss of precision without reversal, whereas Comparator-C identified direction dependence.

MKJ illustrates the distinction. Removing NSA in the later period produced ROR 0.776 (LCL 0.459), while the full-comparator estimate remained positive. Concentration alone did not determine fragility: DQK and DSY had concentrated comparator compositions but remained Comparator-A (Fig. 2D).

Historical persistence

All 19 signals eventually reached two-consecutive cumulative confirmation, and cumulative evidence remained persistent thereafter (Fig. 3A). This common cumulative pattern did not distinguish the robustness classes. It illustrates lock-in: once earlier evidence has accumulated, later local weakening can be obscured by the growing historical denominator.

Rolling windows provided the temporal distinction (Fig. 3B,C). Each of the five most robust signals had LCL > 1 in all 21 windows. The 13 predefined fragile signals had a median persistence of 0.857 (IQR 0.714–0.952); the rank-biserial effect was 0.846 and the two-sided P value was 0.006. One additional two-period signal belonged to neither comparison group and was excluded from the 5-vs-13 exploratory comparison. MKJ retained ROR > 1 in all rolling windows but LCL > 1 in 6/21 windows (Fig. 3D).

Illustrative device phenotypes

DTZ flow/occlusion had an earlier ROR of 3.95 (95% CI 3.57–4.37) and a later ROR of 8.69 (7.71–9.80). LCL > 1 was retained in all eight quarters and all six annual LOO analyses; the signal was Comparator-A and

had rolling LCL > 1 in 21/21 windows. The dominant exact terms included Fluid/Blood Leak, Leak/Splash and Gas Output Problem (Fig. 4A,B).

DWF structural fracture showed earlier and later RORs of 6.43 (95% CI 5.34–7.73) and 4.48 (3.62–5.54), with quarter LCL > 1 in 8/8, annual LOO LCL > 1 in 6/6, Comparator-A status and rolling LCL > 1 in 21/21. Crack, Break, Fracture and Material Rupture formed a coherent structural vocabulary across 2020–2025 (Fig. 4C,D).

LOX structural fracture had smaller broad-period effects but remained Comparator-A and had rolling LCL > 1 in 21/21 windows. MKJ met both broad-period criteria but had quarter LCL > 1 in 2/8, later manufacturer-removal ROR 0.530 after removing ZOLL MEDICAL (95% CI 0.350–0.801), Comparator-C status and rolling LCL > 1 in 6/21. These cases are engineering archetypes, not a separate case series.

Recall context

Legacy high-confidence recall rows collapsed to fewer deduplicated high-concordant events: DTZ 43 to 17, DWF 18 to 5, LOX 1 to 1 and MKJ 84 to 42. Among 207 high-concordant two-period pair-events with valid initiation dates, 163 were initiated before the first replayed statistical crossing, one in the same quarter and 43 after. Recall linkage therefore supplied regulatory context rather than independent validation or prediction (Supplementary Figure S3).

Discussion

The central methodological finding is that the same broad-period positivity can conceal materially different analytical robustness. In this fixed universe, 19 pairs met formal criteria in both broad periods, but only five met the full temporal and entity-label criteria. A broad-period signal should therefore not be treated as a uniformly stable result simply because it passes the same threshold twice.

Temporal-window analysis, entity-label removal and comparator perturbation measured different fragility dimensions. A signal could remain Comparator-A while depending on one submitted manufacturer label, or retain local temporal support while losing direction after a comparator code was removed. These dimensions should be inspected separately rather than collapsed into a single unvalidated score.

Cumulative replay and rolling analysis answered different questions. Cumulative evidence remained persistent for all 19 signals, whereas rolling windows showed whether support continued locally. MKJ made this contrast visible: its rolling ROR stayed above 1, but its LCL exceeded 1 in only 6/21 windows. Historical accumulation can preserve a signal after local precision has weakened.

This stress-testing approach may be useful for methodological practice in spontaneous and post-market reporting, where analytical specification is part of the evidence chain. The present work is an empirical demonstration in FDA MAUDE, not a validated framework, universal method or prediction system. The illustrative phenotypes add engineering context but do not change that methodological scope.

The study has several limitations. The 61-pair universe and upstream refinement were selected before the present stress tests, and the definitions were frozen after review of the complete 2020–2025 record rather than in a truly prospective design. Operational thresholds were not externally validated. MAUDE lacks an exposure denominator and is affected by missing or heterogeneous reporting, possible related MDR submissions and pair-level non-independence. Exact manufacturer and model labels do not necessarily represent corporate or model families. The selected groups and recall comparisons are post-selection, small and non-independent at the pair level. No second independent surveillance database was used, and recall linkage was not a gold standard. External surveillance datasets with exposure or engineering denominators are needed to test transportability and generalizability.

Conclusions

Broad-period consistency is not sufficient evidence of methodological robustness. Temporal, entity-label and comparator perturbations, together with rolling local persistence, provide complementary sensitivity information, while recall linkage remains contextual. This empirical MAUDE demonstration supports testing reporting signals across analytical specifications before treating them as stable findings.

Abbreviations

AED	
automated external defibrillator	
BH	
Benjamini–Hochberg	
CI	
confidence interval	
CPB	
cardiopulmonary bypass	
ECMO	
extracorporeal membrane oxygenation	
FDA	
U.S. Food and Drug Administration	
FDR	
false discovery rate	
HHI	
Herfindahl–Hirschman index	
LCL	
lower confidence limit	
LOCO	
leave one comparator code out	
LOO	
leave one year out	

MAUDE

Manufacturer and User Facility Device Experience

MDR

Medical Device Report

PTCA

percutaneous transluminal coronary angioplasty

ROR

reporting odds ratio.

Declarations

Ethics approval and consent to participate

This study analysed only publicly accessible, de-identified FDA MAUDE and Medical Device Recall records. The investigators did not recruit or contact participants, collect new human data or specimens, or access identifiable individual-level information. Project-specific institutional ethics approval and informed consent were therefore not required.

Consent for publication

Not applicable.

Clinical trial number

Not applicable.

Availability of data and materials

The datasets analysed in this study are publicly available from the U.S. Food and Drug Administration MAUDE/MDR Data Files and Medical Device Recalls resources [1, 2, 15]. Compact derived datasets supporting the reported results are included in Additional file 2, and the analysis/provenance scripts and portable figure generator are included in Additional file 3. Raw FDA files are not redistributed because they remain publicly accessible from the original FDA resources.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Shandong Provincial Natural Science Foundation (grant no. ZR2020MH029 to Hourong Sun). The funder had no role in study design, data collection, analysis, interpretation, manuscript preparation, or the decision to submit.

Authors' contributions

DX conceptualized the study, curated the data, performed analyses and verification, developed code and visualizations, interpreted results and drafted the manuscript. ZY contributed to data curation, verification, interpretation and critical revision. AL, ZL, QC and JL contributed to data curation, verification, interpretation and review. HS supervised the study, administered the project, acquired funding and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. U.S. Food and Drug Administration. About Manufacturer and User Facility Device Experience (MAUDE) Database. FDA. <https://www.fda.gov/medical-devices/mandatory-reporting-requirements-manufacturers-importers-and-device-user-facilities/about-manufacturer-and-user-facility-device-experience-maude-database>. Accessed 12 Aug 2026.
2. U.S. Food and Drug Administration. MDR Data Files. FDA. <https://www.fda.gov/medical-devices/medical-device-reporting-mdr-how-report-medical-device-problems/mdr-data-files>. Accessed 12 Aug 2026.
3. Zangrillo A, Landoni G, Biondi-Zoccai G, Greco M, Greco T, Frati G, et al. A meta-analysis of complications and mortality of extracorporeal membrane oxygenation. *Crit Care Resusc*. 2013;15(3):172–8. 10.1016/S1441-2772(23)01792-1.
4. Figueroa Villalba CA, McMullan DM, Reed RC, Chandler WL. Thrombosis in extracorporeal membrane oxygenation (ECMO) circuits. *ASAIO J*. 2022;68(8):1083–92. 10.1097/MAT.0000000000001605.
5. Leerson J, Tulloh A, Tovar Lopez F, Gregory S, Buscher H, Rosengarten G. Detecting oxygenator thrombosis in ECMO: a review of current techniques and an exploration of future directions. *Semin Thromb Hemost*. 2024;50(2):253–70. 10.1055/s-0043-1772843.
6. Schwarzman LS, Ishaaya EC, Patel D, Megowan N, Thomas JL. Characteristic adverse events with intra-aortic balloon pumps: an analysis of the U.S. Food and Drug Administration MAUDE database from 2016 to 2021. *Cardiovasc Revasc Med*. 2023;56:37–40. 10.1016/j.carrev.2023.05.432.
7. Khalid N, Rogers T, Shlofmitz E, Chen Y, Musallam A, Khan JM, et al. Adverse events and modes of failure related to Impella RP: insights from the Manufacturer and User Facility Device Experience (MAUDE) database. *Cardiovasc Revasc Med*. 2019;20(6):503–6. 10.1016/j.carrev.2019.03.010.
8. Khalid N, Javed H, Rogers T, Hashim H, Shlofmitz E, Chen Y, et al. Adverse events and modes of failure related to the FilterWire EZ Embolic Protection System: lessons learned from an analytic review of the FDA MAUDE database. *Catheter Cardiovasc Interv*. 2019;94(1):157–64. 10.1002/ccd.28297.
9. Case BC, Yerasi C, Forrestal BJ, Kumar S, Musallam A, Chezar-Azerrad C, et al. Real-world experience of the MANTA closure device: insights from the FDA Manufacturer and User Facility Device Experience (MAUDE) database. *Cardiovasc Revasc Med*. 2021;27:63–6. 10.1016/j.carrev.2020.11.023.

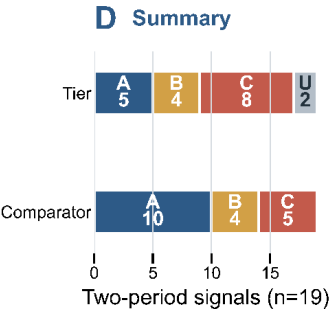
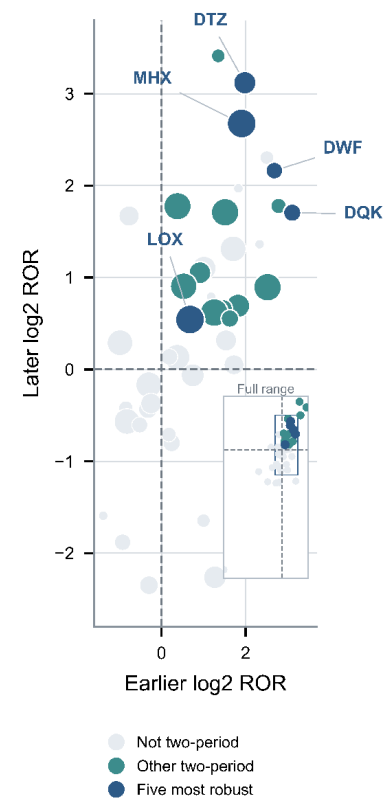
10. Gao Z, Lei W, Gao E, Bhatia S. Device-related patient outcomes for coronary stents: a MAUDE database analysis. *Heliyon*. 2024;10(24):e40908. 10.1016/j.heliyon.2024.e40908.
11. Khraisat O, Manasrah N, Mosaad MA, Aljabali A, Samardali H, Ismail W, et al. Device-specific failure patterns of guide catheter extensions in percutaneous coronary intervention: a contemporary MAUDE database analysis (2022–2025). *Cardiovasc Revasc Med*. 2026. 10.1016/j.carrev.2026.04.022.
12. Chung G, Etter K, Yoo A. Medical device active surveillance of spontaneous reports: a literature review of signal detection methods. *Pharmacoepidemiol Drug Saf*. 2020;29(4):369–79. 10.1002/pds.4980.
13. Blake K. Postmarket surveillance of medical devices: current capabilities and future opportunities. *J Interv Card Electrophysiol*. 2013;36(2):119–27. 10.1007/s10840-013-9778-6.
14. Batheja A, Bentan M, Coelho DH. Considerations for use of the MAUDE database in otolaryngology-head & neck surgery research. *Otolaryngol Head Neck Surg*. 2024;171(1):320–2. 10.1002/ohn.666.
15. U.S. Food and Drug Administration. Medical Device Recalls. FDA.
<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfres/textsearch.cfm>. Accessed 12 Aug 2026.

Figures

A Analytic workflow



B Earlier versus later ROR landscape (61 pairs)



C Six-domain robustness matrix

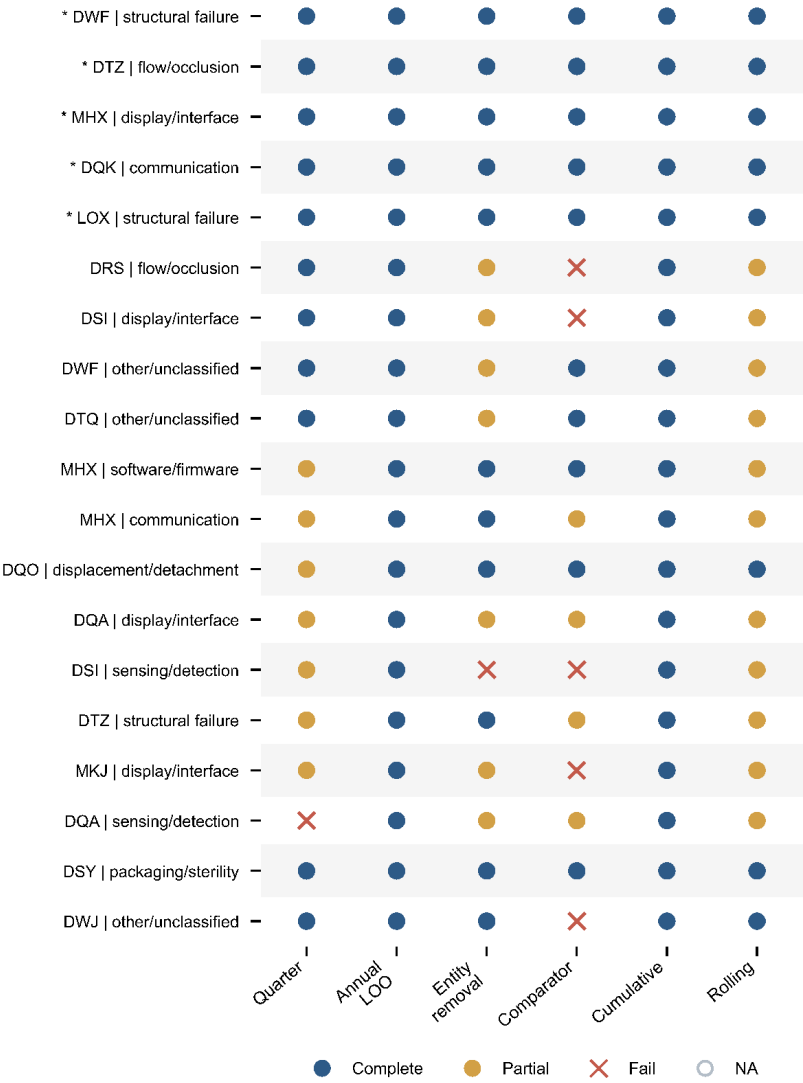


Figure 1

From broad-period consistency to the global signal landscape.

A, Analytic workflow from 2020–2025 FDA MAUDE records to 61 evaluable pairs, 26 earlier-period and 20 later-period formal signals, 19 two-period signals, and three downstream robustness modules: temporal/entity robustness, comparator robustness and historical persistence; recall records are shown as

contextual information only. B, earlier-versus-later log2 RORs for all 61 pairs, with the main plot zoomed to the principal data-dense region and an inset showing the complete range, including extreme points; the five most robust signals are labeled. C, six-domain symbol matrix for the 19 two-period signals, covering quarter support, annual leave-one-year-out support, entity-removal support, comparator robustness, cumulative historical persistence and rolling-window support. D, compact temporal/entity tier and comparator-class summary. LCL, lower confidence limit; LOO, leave one year out; ROR, reporting odds ratio. The plotted classifications are frozen results; ROR is reporting disproportionality, not incidence or comparative safety.

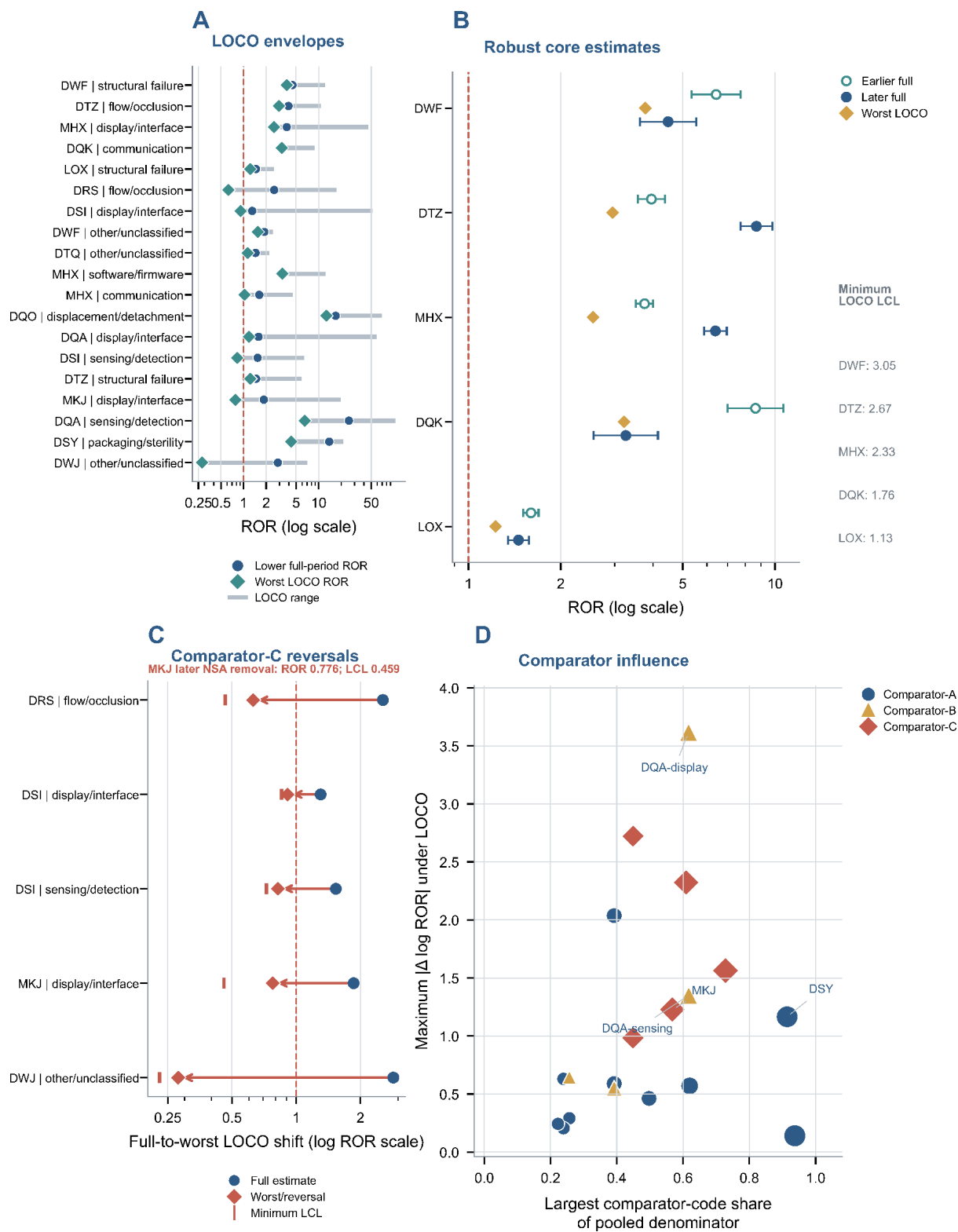


Figure 2

Comparator sensitivity.

A, lower full-period ROR, worst LOCO ROR and the LOCO range across earlier- and later-period scenarios. B, earlier and later full estimates, worst LOCO ROR and minimum LOCO LCL for the five most robust signals. C, five Comparator-C direction-reversal signals, including the MKJ later-period NSA-removal sensitivity

(ROR 0.776; LCL 0.459). D, comparator-code concentration and maximum LOCO influence. LOCO, leave one comparator code out; HHI, Herfindahl–Hirschman index. Comparator concentration is descriptive, not a failure criterion.

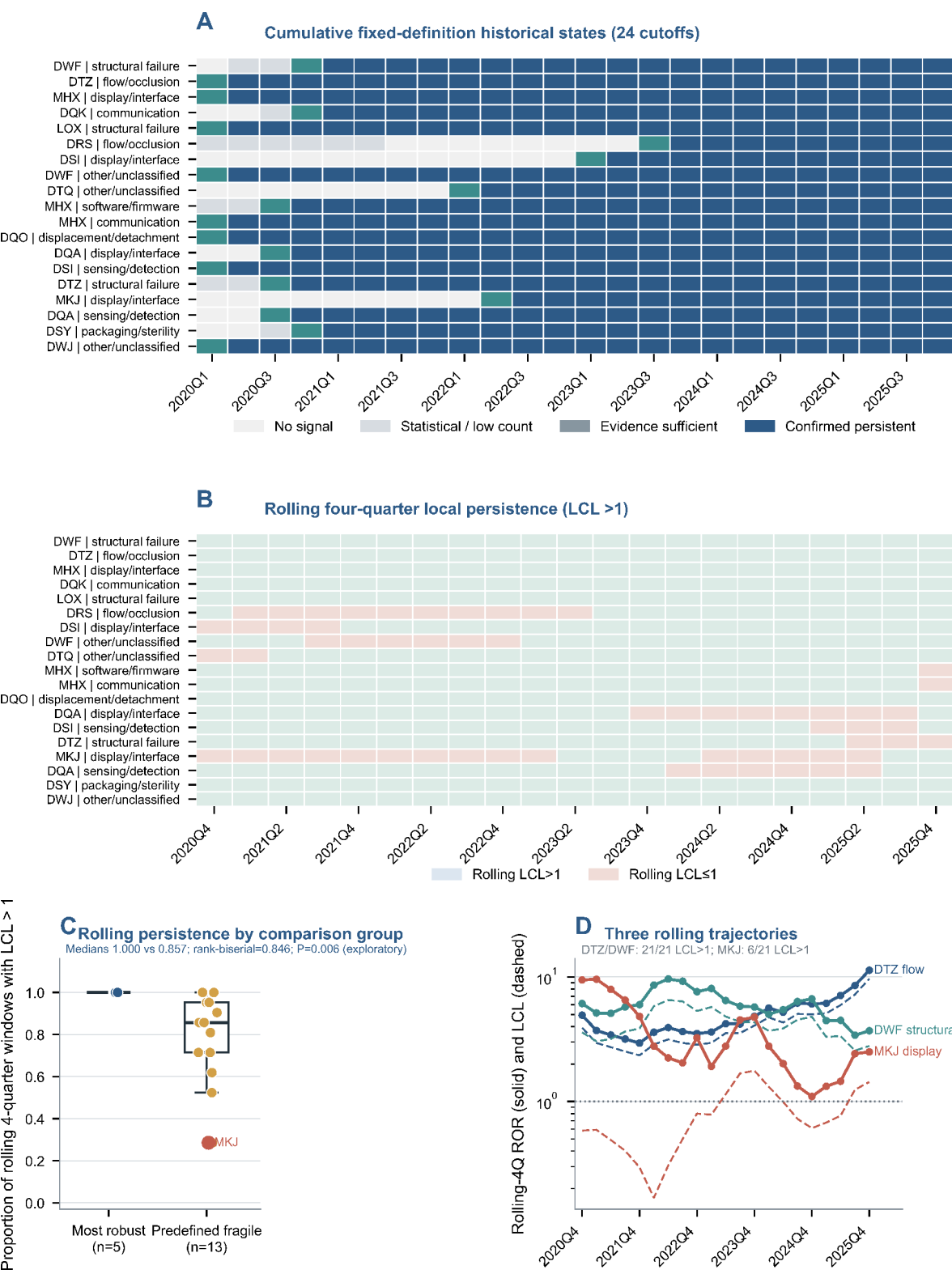
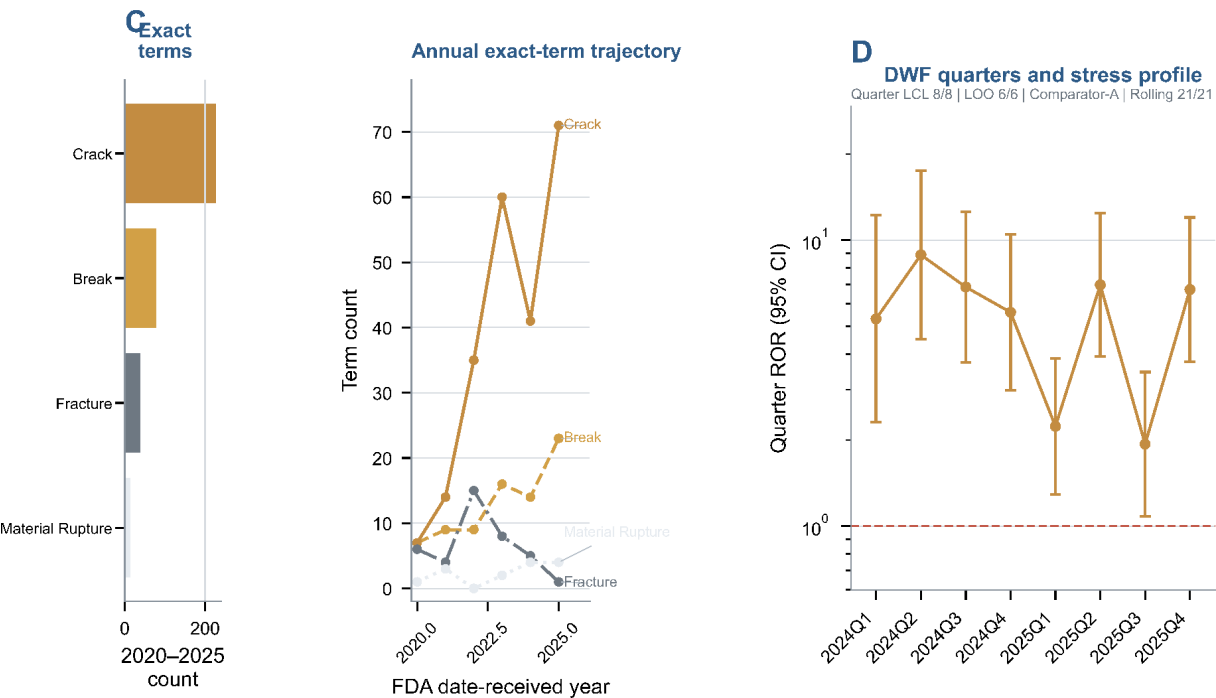
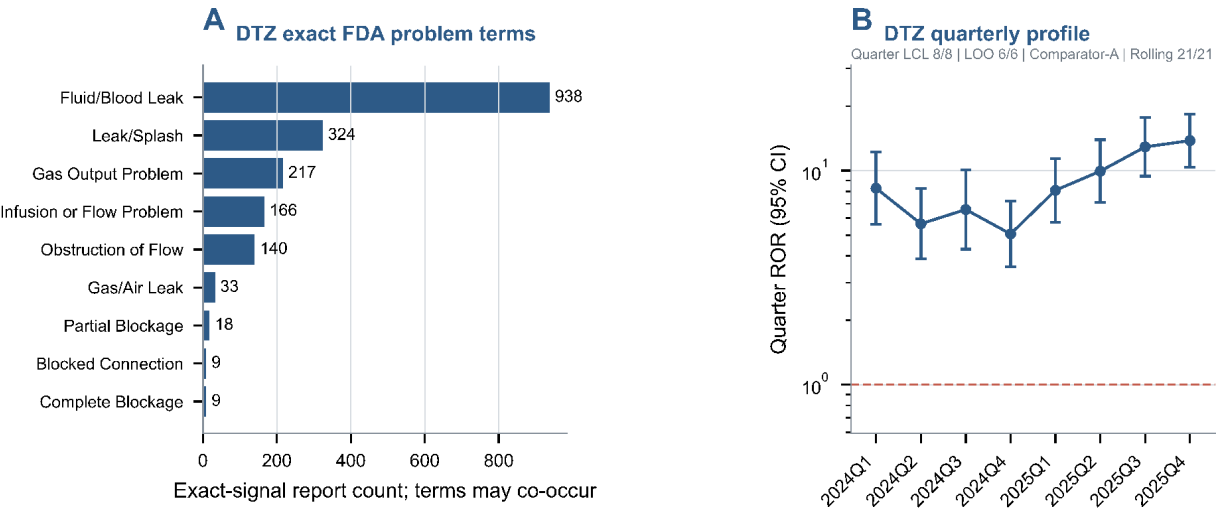


Figure 3

Historical persistence.

A, cumulative states across 24 cutoffs. B, rolling four-quarter LCL > 1 status across 21 windows. C, rolling persistence in the five most robust and 13 predefined fragile signals; one additional two-period signal was excluded from this exploratory comparison; medians were 1.000 versus 0.857 with rank-biserial 0.846 and two-sided P=0.006. D, rolling ROR and LCL trajectories for DTZ, DWF and MKJ. The replay was conditional on definitions frozen after complete 2020–2025 review and was not prospective validation; the group comparison was exploratory, post-selection and affected by pair-level non-independence.



	Quarter persistence	Manufacturer robustness	Comparator robustness	Rolling persistence
LOX structural, lower-effect durable	6/8	2/2	Comparator-A	21/21
MKJ display/interface, fragile	2/8	1/2	Comparator-C	6/21
Manufacturer removal: ZOLL MEDICAL -> ROR 0.530 (95% CI 0.350–0.801)				
LOCO removal of NSA -> ROR 0.776; LCL 0.459				

Figure 4

Illustrative device phenotypes.

A, exact FDA problem terms for DTZ. B, DTZ quarterly RORs. C, DWF exact terms and annual trajectory. D, DWF quarterly RORs. E, four-domain comparison of LOX and MKJ. These cases are illustrative archetypes within the complete analysis universe; exact FDA problem terms may co-occur within an MDR report and do not establish device causation.

Supplementary Files

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

- [Additionalfileslegends.docx](#)
- [Additionalfile1.pdf](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile3.zip](#)