Research Square

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

Historical Evaluation of Integrated Project Control Early Warning Indicators in Four United States Transit Capital Projects

Mohammed Adeyemi Yusuff

Yemiyusuf191@gmail.com

Northeastern University

Research Article

Keywords: Project controls, Early-warning indicators, Historical back-testing, Public transit, Schedule risk, Capital project delivery

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11283302/v1

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

Additional Declarations: No competing interests reported.

Title: Historical Evaluation of Integrated Project Control Early Warning Indicators in Four United States Transit Capital Projects

Mohammed Adeyemi Yusuff

College of Professional Studies, Roux Instiute, Northeastern University, Portland, Maine, USA.

ORCID: 0009-0000-2801-1308

Yemiyusuf191@gmail.com

Abstract

Major U.S. transit capital projects are monitored through cost and schedule measures that may not fully represent change, contractor and construction-interface pressures. This study retrospectively applied the frozen Integrated Project Controls and Risk Forecasting Framework (IPCRF), Version 1.2, to four U.S. transit capital projects with contrasting outcomes: NYCT ADA station improvements, Dulles Corridor Metrorail Phase 2, East Side Access/Grand Central Madison, and the Canarsie Tube/L Project Tunnel Rehabilitation as a favorable-outcome comparator. Twelve selected project-level historical observations were reconstructed from transit-agency and Federal Transit Administration records; Case 4 also contributed eight component-level contingency-stress observations. Additional monthly PMOC records for Dulles and East Side Access were examined in a reporting-period sensitivity analysis. NYCT Finish Slippage moved from Low to Critical after its completion forecast changed but before the original completion target. Dulles retained Low project-level EAC variance while Total Change Exposure remained High, alongside documented contractual milestone failures. East Side Access showed Elevated cost and conditional schedule classifications with access and interface constraints. Canarsie remained Low for Finish Slippage at both historical snapshots; seven component Contingency Stress observations were Low, and one was Watch, with none Elevated or above. Incomplete and uneven records prevented formal six-domain PDRS/FCS scoring. The findings support retrospective indicator-level reconstruction across contrasting outcomes, but do not establish specificity, false-positive performance or prospective predictive validity.

Keywords: *Project controls; Early-warning indicators; Historical back-testing; Public transit; Schedule risk; Capital project delivery.*

1. Introduction

1.1. Project delivery challenges in U.S. transit capital projects

Delivering major U.S. transit capital projects has remained a significant challenge due to cost growth, schedule slippage and construction coordination issues. These projects involve substantial financial commitments and long implementation periods, as well as interdependent construction

activities that can make it difficult to achieve the approved delivery objectives. A study of 81 U.S. rail transit projects by Gao and Touran (2020) confirms that cost overruns increase with larger projects, and the more time that passes, the worse the estimates become. Evidence from other U.S. transportation infrastructure projects also shows that cost, change-order and schedule performance do not necessarily improve in tandem. Shrestha et al. (2022) analyzed 3,957 highway projects in Texas to find cost growth reductions with no corresponding improvements in schedule performance. Although highway projects have different characteristics from passenger rail projects, the following statistics underscore the need to measure different delivery outcomes and not equate improvement in one metric with general project performance.

Reliable project control information is important because of evidence of federal oversight. In regard to capital cost estimates of past federally supported transit projects, the U.S. Government Accountability Office (GAO, 2023) noted that capital cost estimates for previously completed projects have improved, yet also highlighted ongoing issues with determining project outcomes. There are also varying delivery and monitoring needs that differ for each individual transit project. Changes in completion forecasting and construction conditions are noted on the accessibility improvement records, produced by the Metropolitan Transportation Authority (MTA, 2023b). The Washington Metropolitan Area Transit Authority (WMATA, 2022) reported that operation of the Dulles Silver Line extension began in November 2022. These records may be used to assess if historical project control information contained any observable signs of deterioration in the delivery process before the subsequent outcomes.

1.2. Limitations of conventional project-control approaches and the evidence gap

Conventional project-control measures may not fully capture emerging delivery stress when cost, schedule, contractor performance and construction interfaces are interpreted separately. The Federal Transit Administration (FTA, 2024) recognises the interdependence of project scope, cost, schedule, risk and contingency in its capital project oversight procedures. However, monitoring several dimensions does not itself establish how their warning behaviour changes or converges across reporting periods. Empirical evidence from urban rail projects in Vietnam identified design changes, contractor capability, project resources and financing among factors affecting both cost and schedule outcomes (Nguyen et al., 2024). Likewise, Muruganandan et al. (2022) showed, with the Crossrail project, that evolving interfaces and interdependencies demand continuous coordination and control through construction systems' maturation.

Further research on early-warning approaches justifies the examination of multiple project-control signals. Cho et al. (2021) identified combinations of early project conditions associated with subsequent schedule delays across 19 precast-concrete projects in Korea. Meanwhile, Shafaat et al. (2023) developed change-order warning models using historical data from 5,628 US highway projects, and Lagos et al. (2024) demonstrated the value of combining production-planning information with conventional schedule indicators for forecasting construction performance. González-Cruz et al. (2022) further proposed a schedule-oriented index that could predict project

delays. The studies confirm the relevance of advance indicators, but derive from various contexts of infrastructure, information structure and analysis. They do not establish whether IPCRF indicators can be reconstructed consistently from historical U.S. transit oversight records or what additional information their joint interpretation provides beyond existing headline measures. A further methodological concern is outcome-selection imbalance: a retrospective back-test based only on projects with documented delivery problems cannot show whether the same fixed rules remain comparatively quiet when applied to a project with a documented favorable outcome.

1.3. Study rationale, aim and research questions

Historical backtesting offers a chance to assess the performance of project control indicators, based on conditions which represent a known historical set of projects. Unlike final deliverables-based assessments, reconstructing the different reporting periods provides insight into when warning signs were first detected, and how that classification differed from the indications in earlier milestones. This type of temporal evaluation is critical because construction schedule risks may arise when activity conditions and task relationships change before the risks can be realized. Zhang and Wang (2023) demonstrated how data-driven reliability analysis can be used to identify critical construction tasks and evaluate schedule-delay risk dynamically. Applying an indicator framework, however, is retrospective and requires differentiation of historical information and later known results.

This study examines Version 1.2 of the Integrated Project Controls and Risk Forecasting Framework (IPCRF), developed by Yusuf (2026), using historical records from four U.S. transit capital projects: NYCT ADA station improvements; Dulles Corridor Metrorail Phase 2; East Side Access/Grand Central Madison; and the Canarsie Tube/L Project Tunnel Rehabilitation. The first three cases provide deterioration-enriched historical evidence, while Canarsie provides a documented favorable-outcome comparator. Four research questions are addressed: (1) what project-control conditions are identified by reconstructable IPCRF indicators across cases with contrasting outcomes; (2) where deterioration occurred, which indicators produced the earliest observed qualifying classifications within the reviewed records, and whether the favorable comparator remained below severe thresholds; (3) whether multi-domain information revealed conditions not apparent from headline cost or finish-date measures alone; and (4) how data completeness, rebaselining, reporting structure and measurement compatibility affected reconstruction and interpretation.

The study therefore provides an exploratory retrospective historical back-test of indicator behaviour across four U.S. transit capital projects. Because Version 1.2 was applied after the principal historical outcomes had occurred, the analysis does not constitute prospective predictive validation. Incomplete quantitative coverage also prevents formal assessment of the complete six-domain PDRS and FCS. The study instead examines whether the frozen indicators can reconstruct different historical project-control conditions without changing thresholds or filling missing fields.

2. IPCRF framework and analytical expectations

2.1. Framework architecture and early-warning indicators

The Integrated Project Controls and Risk Forecasting Framework (IPCRF) brings together six project control areas to analyse the impact of financial, scheduling, risk, change, contractor and constructability factors on project delivery. These domains address the interconnected aspects of capital project performance which can deteriorate at various stages of implementation. The Federal Transit Administration (FTA, 2024) also views cost, schedule, and constructability as different parts of the capital project oversight process all linked together. Furthermore, experiences from the Crossrail project in London highlight the need for ongoing coordination when the technical interfaces and interdependencies change over time as they become more integrated throughout operation (Muruganandan et al., 2022).

These relationships are implemented through 23 indicators in IPCRF Version 1.2, including cost, schedule, risk, change, contractor performance and constructability/sequencing criteria (Yusuf, 2026). Cost indicators assess forecast expenditure, cost efficiency, and contingency consumption, while schedule indicators examine completion forecasts, milestone reliability, and float erosion. Risk and change indicators reflect unaddressed exposures and management actions. The delivery commitments and reporting compliance to the project are assessed through contractor performance indicators, while access restrictions, critical issues, and interface failures are analyzed through the constructability indicators.

The framework distinguishes between observed delivery conditions and subsequent outcomes by using indicator trajectories and rule-based warning classifications. It is expected that deterioration in upstream project control conditions will become visible prior to the results in the “headline” measures and costs. This notion, however, must be evaluated empirically: multiple markers are not necessarily associated with early detection, and the indications themselves are not necessarily new markers beyond those already reported in the respective projects.

2.2. Frozen specification, scoring and data-quality requirements

The historical backtest uses the definitions and thresholds of the IPCRF Version 1.2 indicator without recalibrating them to the observed project outcomes. Individual values are assigned to one of five categories of increasing sensitivity (Low, Watch, Elevated, High and Critical). The available indicator scores are then combined using predefined weights within each domain, while the Project Delivery Risk Score (PDRS) assigns a weighting of 20% to cost, 25% to schedule, 20% to risk, 15% to change, and 10% each to contractor performance and constructability/sequencing. The framework also specifies trend persistence and compound warning rules intended to distinguish isolated deterioration from concurrent multi-domain stress.

Version 1.2 permits redistribution of unavailable indicator weights within a domain, but a domain is not formally scored when more than 25% of its required fields are missing; equivalently, at least 75% completeness is required. Missing or incompatible observations remain unavailable and are not assigned zero-risk values. The present analysis therefore distinguishes independently

reconstructable indicator classifications and diagnostic calculations from formally eligible domain, PDRS and FCS outputs. Redistribution among available indicators does not override the completeness gate.

The September 2026 specification of the framework was held unchanged for the present retrospective reanalysis. This does not prove that the thresholds were set before the historical outcomes of the projects occurred. The evaluation, then, is to inquire about the behavior of indicators under a prescribed retrospective specification and not about prospective predictive validation. In line with the distinction between project predictions and realised capital-project outcomes made by the Government Accountability Office (GAO, 2023), later delivery records are used exclusively as comparison endpoints. The sequence of analysis shown in Figure 1 is a path from historical observations through indicator classification and completeness assessment to observed warning signs, finally to the delivery outcomes.

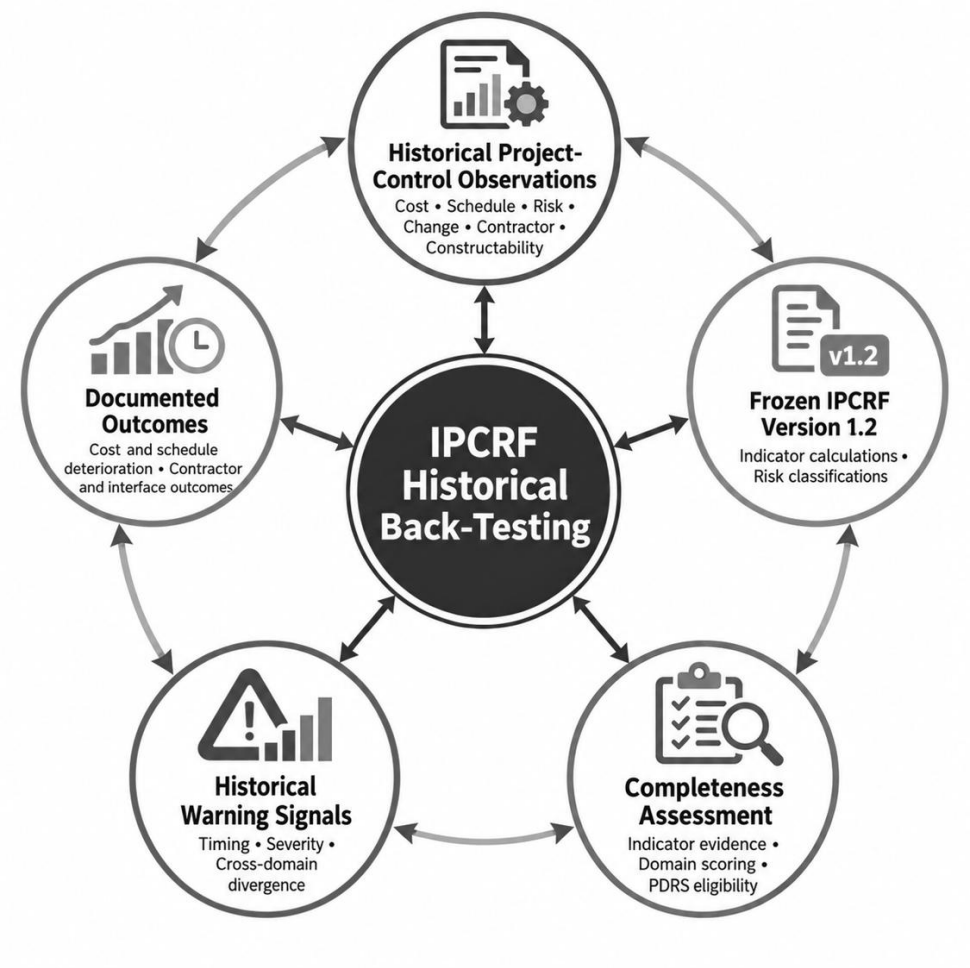


Figure 1. Conceptual and analytical architecture of IPCRF historical back-testing.

Note. IPCRF = Integrated Project Controls and Risk Forecasting Framework; PDRS = Project Delivery Risk Score. The figure illustrates the analytical relationship between historical project-control observations, the fixed framework specification, data completeness, warning signals and subsequently documented project outcomes.

3. Materials and methods

3.1. Study design and analytical protocol

This study used an exploratory, retrospective, longitudinal, multi-case design to examine the historical performance of early-warning indicators within the Integrated Project Controls and Risk Forecasting Framework (IPCRF). This design allowed project control conditions to be reconstructed at successive historical reporting periods, enabling the resulting warning classifications to be compared with subsequently documented delivery outcomes. A multiple-case approach was used to analyze indicator behaviour, applying to various projects, which differed in delivery modalities, reporting systems, and baseline characteristics (Yin, 2018).

The project-period was the primary analytical unit, with each project-level observation representing information available at a defined historical reporting point. IPCRF Version 1.2 was held frozen throughout the analysis; its indicator definitions, formulas, thresholds and weights were not altered in response to the known outcomes. Because the framework was applied after the principal project outcomes had occurred, the study is described as retrospective historical back-testing rather than prospective validation. Historical inputs and later outcomes were kept separate, and no synthetic observations were introduced. The analytical protocol comprised case selection, source verification, historical extraction, indicator reconstruction, classification, completeness assessment, outcome comparison and cross-case synthesis.

3.2. Case eligibility, selection and project characteristics

Four U.S. transit capital projects were examined: NYCT ADA station improvements at 149 St–Grand Concourse and Tremont Av; Dulles Corridor Metrorail Phase 2; East Side Access/Grand Central Madison; and the Canarsie Tube/L Project Tunnel Rehabilitation. Selection remained purposive and evidence-based rather than statistically representative. The three original cases were retained because their historical records allowed reconstruction of selected project-control conditions associated with documented delivery deterioration.

Canarsie was added as a purposive favorable-outcome comparator rather than as another deterioration-enriched case. Its public record provides two project-level point-in-time schedule snapshots, component-level cost and schedule control information, and a documented favorable completion outcome. The case was therefore used to examine whether the same frozen Version 1.2 rules remained comparatively quiet when applied to a project that was later documented as completed ahead of schedule and under budget. Its more limited domain coverage was retained explicitly rather than supplemented with synthetic or non-equivalent measurements.

The resulting four-case sample supports comparison of historical indicator behaviour across contrasting delivery conditions. It does not support population-level estimates of sensitivity, specificity or discriminatory accuracy. Table 1 summarises project scope, selected reporting periods, baseline references, reconstructable domains and subsequent outcomes.

Table 1. Characteristics, historical reporting periods, baseline references and analytical coverage of the included U.S. transit capital projects

Project and scope	Selected historical reporting periods	Baseline budget and schedule reference	Available project-control domains	Subsequent outcome reference
NYCT ADA Station Improvements: 149 St–Grand Concourse Complex and Tremont Av; design–build accessibility and state-of-good-repair improvements	July 2021; January 2022; February 2023; October 2024	Historical budget reference: $160.409 million; original completion target: 1 July 2023	Schedule: Finish Slippage. Cost: limited, non-continuous budget/EAC observations	Later MTA records document repeated forecast changes and an estimated completion of 30 June 2026; verified project-wide actual completion was not established in the reviewed evidence
Dulles Corridor Metrorail Phase 2: Wiehle Avenue to Ashburn; Silver Line extension	March, June and December-labelled 2019 records	Project-level cost basis: $2,778.236 million; accepted rebaselined RSD: 31 March 2020	Cost, Schedule, Change and selected Contractor Performance evidence; Risk and Constructability evidence incomplete for formal scoring	Revenue service commenced 15 November 2022, 959 days after the accepted baseline
East Side Access / Grand Central Madison: commuter-rail terminal and associated infrastructure connecting LIRR services to Grand Central	October 2018; May 2019; September 2019	June 2014 re-plan budget: $10,177.8 million; internal target RSD: 12 February 2021; public RSD: 13 December 2022	Cost and Schedule indicators; selected Constructability/Sequencing diagnostics; other domains incomplete	Limited shuttle service commenced 25 January 2023; full service commenced 27 February 2023, 76 days after the public RSD
Canarsie Tube / L Project Tunnel Rehabilitation: New York City Transit	29 September 2019; 25	Schedule comparison used the	Schedule: Finish Slippage. Cost: component-level Contingency Stress mapping.	Tunnel work completed 26 April 2020; later MTA material reports

tunnel rehabilitation and associated works	March 2020; 2Q2020 component-control snapshot reported separately	source-stated prior planned finish. Public cost figures were retained on their reported scope bases; no compatible project-level BAC/EAC series was constructed	MTA cost-index and schedule-variance values retained as context only	completion three months ahead of schedule and $100 million under budget

Note. The four-case primary project-level dataset contains 12 historical observations. The Canarsie 2Q2020 Traffic Light data are component-level observations and are therefore reported separately from the 12-period project-level denominator. Subsequent outcomes were used only for retrospective comparison and were not incorporated into historical warning calculations. Available domains indicate reconstructable evidence and do not imply formal domain-scoring eligibility.

3.3. Historical data sources and extraction procedures

Historical observations were compiled by reconstructing them from official monitoring reports by the Federal Transit Administration (FTA), Capital Program Committee (CPC) records of the Metropolitan Transportation Authority (MTA), and from transit agency project documentation. For NYCT the July 2021, January 2022, February 2023, and October 2024 reports included project-level schedule forecasts, budget information and documented delivery conditions (MTA, 2021; 2022; 2023b; 2024).

Dulles observations were extracted from the March, June and December 2019 PMOC reports, which included project cost estimates, earned value, actual expenditure, contingencies, approved and pending changes, contractual milestones, and schedule forecasts (FTA, 2019a–c). East Side Access observations were obtained from the October 2018, May 2019 and September 2019 PMOC reports. These historical baseline conditions (2014, 2018, 2019d, 2019e) were based on the report from December 2014.

Canarsie evidence was reconstructed from official FTA and MTA records preserved in the Case Study 4 evidence package (DOI: 10.5281/zenodo.23088970). Project-level schedule snapshots were retained for 29 September 2019 and 25 March 2020. The MTA 2Q2020 Traffic Light Report supplied eight component-level contingency indices together with component cost indices and quarter-over-quarter schedule-variance values. The 26 April 2020 completion record and later MTA statement that the tunnel rehabilitation finished three months ahead of schedule and $100 million under budget were retained as outcome checks. Component-, contract-, tunnel- and broader

programme-level cost figures were kept separate where a common scope basis could not be demonstrated.

A harmonised extraction register, including project identifiers, source references, reporting cut-off dates, baseline data, current forecasts, inputs to indicators and completeness indicators, was created. Original and revised baselines were retained separately. Extracted values were verified with supporting narratives, or against their sources in the tables, paying particular care to units of measures, project scope and whether the reported value concerned forecasts or outcomes.

Differing reporting periods, underlying data cut-off dates and document finalisation dates were logged separately. The first public-release dates were only retained where they were established independently; the finalisation of reports was not assumed to demonstrate public availability. This is important to ensure that a warning date is not set before supporting evidence can be provided. Original project records and harmonised extraction register offer an audit trail to follow when subsequent calculations are performed. Supplementary Table S2 presents the source verification and temporal provenance register, including the reporting periods, baseline definitions, and historical outcome references used in the analysis.

3.3.1. Sensitivity to reporting-period selection

The primary harmonised project-level dataset contained 12 historical observations: four for NYCT ADA station improvements, three for Dulles Phase 2, three for East Side Access and two for the Canarsie Tube/L Project. Additional consecutive PMOC records were available for Dulles and East Side Access and were reviewed separately in the reporting-period sensitivity analysis. The eight Canarsie 2Q2020 component records were retained as component-level evidence rather than converted into an additional project-level observation.

This analysis focused on indicators that contributed directly to the main findings rather than reconstructing all 23 IPCRF indicators for each monthly report. For each compatible record, the analysis examined whether an omitted reporting period had identified an earlier 'Elevated', 'High' or 'Critical' classification; whether it had changed the maximum observed severity; whether it had demonstrated persistence of the classification; or whether it had revealed an intervening recovery. Measurements lacking compatible definitions, denominators, or baseline information were not imputed, but remained unavailable.

The additional Dulles and East Side Access records remained a secondary sensitivity dataset and were not added to the 12 primary project-level observations. No separate monthly sensitivity series was constructed for Canarsie. Its two historical schedule snapshots, component-level 2Q2020 controls and outcome-state checks, were retained according to their original analytical level.

3.4. Indicator reconstruction and scoring procedures

The project control indicators available were computed from the definition and thresholds provided in IPCRF Version 1.2 (Yusuf, 2026). The framework comprises six domains: Cost,

Schedule, Risk, Change, Contractor Performance and Constructability/Sequencing. Historical variables were only mapped to the appropriate indicators if the source information met the measurement definition and project level basis required.

Cost calculations included estimate-at-completion (EAC) variance, cost performance index (CPI), contingency stress, if applicable. Schedule calculations used foresee slippage of forecasts and recorded milestone performance. Records of change exposure were used to infer change and contractor performance indicators. The constructability evidence was then compared to the definitions of access conflict and interface failure as defined by the framework. Published narrative descriptions were not automatically converted into quantitative measurements in the absence of defensible event identification and denominators (Yusuf, 2026; FTA, 2019e).

For example, the East Side Access programme contingency depletion measure was not included in the concurrent critical path method measure, because while the reported quantities are part of the same measure, they reflect scheduling constructs rather than depletion. Similarly, selected coordination-point delays were distinguished from a verified project-wide interface-failure rate. Supplementary Table S3 details the underlying East Side Access coordination-point activities and event level classifications.

The primary calculations were:

$$\text{EAC Variance (\%)} = \frac{EAC - BAC}{BAC} \times 100$$

$$\text{Finish Slippage (\%)} = \frac{\max(0, F_c - F_b)}{D_b} \times 100$$

Where 'BAC' denotes the applicable baseline budget, 'F_c' denotes the contemporaneous forecast finish, 'F_b' denotes the specified baseline finish and 'D_b' denotes the baseline duration in days. Where construction start dates were estimated, the resulting normalised values were explicitly identified as proxy-dependent. Calculable indicators were assigned their unchanged Version 1.2 risk classifications. The formulas, provisional threshold bands and within-domain weights for all indicator types contributing to the retained primary classifications are presented in Supplementary Table S4 (Yusuf, 2026). Missing indicators were retained as unavailable rather than assigned zero-risk scores. Under Version 1.2, a domain was not formally scored when more than 25% of its required fields were missing; at least 75% completeness was therefore required. Diagnostic calculations were kept separate from formally eligible domain scores, and neither a complete PDRS nor FCS was reported where the completeness gate could not be demonstrated. Component-level Canarsie observations were not aggregated into project-level domain scores because Version 1.2 provides no rule for doing so. The complete register of reconstructed numerical indicators, risk classifications, and measurement qualifications is presented in Supplementary Table S1. Table 2 summarises indicator availability, calculation status and domain-scoring eligibility across the four cases.

Table 2: Availability of IPCRF indicators and domain-scoring eligibility across the four historical transit capital projects

IPCRF domain	NYCT ADA improvements	Dulles Corridor Metrorail Phase 2	East Side Access / Grand Central Madison	Canarsie Tube / L Project	Formal scoring eligibility
Cost	Limited EAC and budget observations; no consistent longitudinal cost series.	EAC Variance, CPI and Contingency Stress available; cost-trend information requires additional consistency checks.	EAC Variance available using the specified 2014 budget and 2018 cost forecast; other indicators unavailable.	Eight component-level Contingency Stress mappings were available: seven Low and one Watch. MTA Cost Index values of 0.49–1.00 were retained as context only. A same-scope project-level BAC/EAC series was unavailable.	**Dulles:** conditional, pending field-level completeness verification. **NYCT and East Side Access:** formal scores withheld. **Canarsie:** component-level evidence only; no project-level cost-domain score calculated.
Schedule	Finish Slippage available; Milestone Delay Rate and Float Erosion unavailable.	Finish Slippage and Milestone Delay Rate available; Float Erosion unavailable.	Finish Slippage available; programme-contingency depletion retained as a diagnostic proxy rather than certified CPM Float Erosion.	Finish Slippage was reconstructable at 29 September 2019 and 25 March 2020; both observations were 0.00%, Low (10). MTA 2Q2020 schedule variance was 0 months for all eight listed components but was retained as context only.	**Indicator-level evidence available in all four cases; formal domain-score eligibility not established.**
Risk	Quantitative risk-register inputs unavailable.	Risk narratives available, but	Risk narratives available, but compatible	Structured probability–impact,	**Not scored in any case.**

		consistent probability–impact and mitigation-rate inputs unavailable.	quantitative inputs insufficient.	mitigation and open-risk series unavailable.	
Change	Approved/pending change values and aging series insufficient.	Total Change Exposure, Pending Change Exposure and Change Trend Growth available at selected periods; Pending Change Aging unavailable.	Change-order narratives available, but consistent exposure and aging series insufficient.	No auditable longitudinal approved/pending change series on a stable scope basis.	**Dulles:** conditional diagnostic scoring; formal eligibility requires field-level verification. **Other cases:** not scored.
Contractor Performance	Contractor-related narratives available; required quantitative rates unavailable.	Missed Contractor Milestones available; update-compliance, late-submittal/RFI and recovery-plan indicators insufficiently quantified.	Contractor delays and submittal concerns documented; complete indicator denominators unavailable.	Required denominator-based milestone, update-compliance, RFI/submittal and recovery-plan series unavailable.	**Dulles:** indicator-level scoring only. **Other cases:** not scored.
Constructability / Sequencing	Access and sequencing concerns documented narratively; quantitative indicator inputs insufficient.	Construction and interface concerns documented; standardised indicator counts/rates unavailable.	Selected Access/Outage Conflicts and Interface Failure events reconstructable; project-wide indicator completeness not established.	Revised construction approach and access constraints documented narratively; standardised issue/interface counts matching Version 1.2 unavailable.	**East Side Access:** diagnostic evidence only; formal domain score withheld. **Other cases:** not scored.

Overall PDRS/FCS	Insufficient formal domain coverage.	Insufficient formal domain coverage despite broader quantitative evidence.	Insufficient formal domain coverage.	Insufficient formal domain coverage; neither PDRS nor FCS calculated.	**Neither complete PDRS nor FCS was calculated for any case.**

Note. IPCRF = Integrated Project Controls and Risk Forecasting Framework; ADA = Americans with Disabilities Act; NYCT = New York City Transit; BAC = budget at completion; EAC = estimate at completion; CPI = cost performance index; CPM = critical path method; RFI = request for information; PDRS = Project Delivery Risk Score; FCS = Forecast Confidence Score. Available indicators are those for which the historical records provided sufficient information for reconstruction, subject to the stated measurement and scope qualifications. Diagnostic calculations are distinguished from formal domain scores. Under IPCRF Version 1.2, a domain is not formally scored when more than 25% of its required fields are missing; equivalently, at least 75% completeness is required. Missing observations were not assigned zero-risk scores. The East Side Access programme-contingency and selected interface-event measurements were retained as diagnostic evidence where equivalence to formal indicator definitions could not be established. For Canarsie, the eight Contingency Stress observations are component-level mappings and were not aggregated into a project-level Cost-domain score. The MTA Cost Index and schedule-variance measures were retained as contextual evidence because their definitions are not identical to the corresponding IPCRF indicators.

3.5. Outcome definitions and warning lead-time analysis

Historical warning signals were contrasted to the predefined project milestones and achieved outcomes. Relevant endpoints included changes in completion forecast, missed contractual milestones, original or approved completion targets and when revenue services started. Forecast deterioration was separated out from realised completion or service outcomes, to prevent it being perceived as a real or verified achievement.

The original July 2023 projected completion date from NYCT served as the key schedule marker, while future dates were deemed as signs of a decline in the overall project (MTA, 2023b, 2024). Dulles was compared to its accepted revenue services baseline, defined as March 2020, and to the documented date that the building opened for business on 15 November 2022 (FTA, 2019a–c; WMATA, 2022). East Side Access retained the 12 February 2021 internal target and the 13 December 2022 late/replanned RSD benchmark as separate schedule references, while limited shuttle service and full service were treated as distinct outcomes (FTA, 2014; MTA, 2023a, 2023c).

Descriptive temporal intervals were calculated as $L = t_o - t_q$, where t_o represents the relevant benchmark or outcome date and t_q represents the earliest compatible qualifying classification identified within the reviewed primary and sensitivity records. Positive intervals indicate that the classification preceded the selected endpoint. These intervals were reported descriptively and were not interpreted as validated early-warning lead times. Where the earliest reviewed observation was already Elevated, High or Critical, the series was treated as left-censored for onset. The analysis therefore identifies the earliest observed qualifying classification supported by the reviewed records rather than the true historical onset of deterioration or first public disclosure.

For Canarsie, the September 2019 and March 2020 schedule classifications were compared with the source-stated planned completion position, while the April 2020 completion announcement was retained only as an outcome check. The later statement that the tunnel rehabilitation was $100 million under budget was not converted into an exact project-level EAC Variance percentage because the matching same-scope budget denominator was unavailable.

3.6. Longitudinal and cross-case analysis

A descriptive, longitudinal analysis of indicator values and risk classifications was undertaken to explore variations in values and classifications, according to the selected reporting periods. Within-case comparisons were conducted to determine if the financial, scheduling, change, contractor, and constructability information suggested convergent or divergent delivery conditions. Particular attention was given to cases in which the headline cost measure remained stable, despite deterioration being documented in other project control dimensions.

Cross-case synthesis was used to compare the earliest observed qualifying classifications, classification severity, descriptive temporal intervals, and the availability of supporting quantitative evidence. Original, revised, internal and later schedule benchmarks were retained separately, and incompatible project scopes or baseline definitions were not aggregated.

The analysis also distinguished between contractual failures that had already occurred and indicators that preceded later programme-level outcomes. Comparisons were restricted to measurements with defensible source values, definitions, and denominators. Because case selection was purposive, the sample was small, domain coverage was unequal and only one favorable-outcome comparator was available, the analysis remained descriptive. No sensitivity, specificity, discrimination statistic or pooled estimate of predictive performance was calculated.

3.7. Data quality, reproducibility and analytical limitations

The quality of the data was assessed through source traceability, temporal consistency, baseline compatibility, indicator completeness, and measurement validity. The supporting source values, formula, reporting reference and evidence status of each calculable indicator were retained. Missing, incompatible and proxy-dependent observations were flagged rather than imputed.

Reproducibility is supported by the reconciled calculation register (see Supplementary Table S1), the source verification and baseline register (see Supplementary Table S2), the East Side Access coordination point register (see Supplementary Table S3), the indicator formula and threshold register (see Supplementary Table S4), and the reporting period sensitivity register (see Supplementary Table S5). However, interpretation is restricted to the reviewed historical evidence and does not establish prospective validity, causation, specificity, or the performance of false positives or negatives at a population level.

4. Results

4.1. Historical dataset characteristics and indicator coverage

The primary harmonised dataset comprised 12 primary project-level observations selected from the following categories: four for NYCT ADA station improvements, three for Dulles Phase 2, three for East Side Access and two for the Canarsie Tube/L Project. Across the 23 IPCRF indicators, these observations represented 276 potential project-level indicator-period cells. The reconciled project-level register contained 49 candidate entries: 20 source-matched indicators, 13 conditional indicators, seven selected-population diagnostics, eight descriptive measurements and one withdrawn calculation. Thirty-three project-level indicator classifications were retained for graphical comparison. Case 4 additionally contributed eight component-level Contingency Stress mappings, comprising seven Low and one Watch classifications. Its 2Q2020 component Cost Index and schedule-variance measures were retained as contextual evidence rather than converted into non-equivalent IPCRF scores. Dulles continued to provide the broadest multi-domain quantitative coverage, while East Side Access and NYCT provided more restricted indicator-level evidence. Canarsie added a favorable-outcome comparator with reconstructable schedule evidence and component-level cost-control information, but insufficient compatible data for formal project-level Cost scoring or the remaining domains. No case provided sufficient compatible coverage for calculation of a complete six-domain PDRS or FCS.

4.1.1. Sensitivity to reporting-period selection

A review of additional compatible monthly PMOC records did not alter the severity or direction of the main cross-case findings significantly, but it did identify earlier qualifying observations for several indicators. For Dulles, total change exposure was already high in February 2019, at 11.57%, preceding the primary observation in March of 11.62%. The February PMOC record's exact-date schedule status showed 148 days of slippage relative to the accepted 31 March 2020 revenue-service baseline. This equates to 6.02% using the same 2,458-day duration proxy, and therefore remains within the elevated band. Pending Change Exposure first exceeded the elevated threshold in April 2019, at around 3.04%, rather than in June.

East Side Access showed a similar timing effect. The May 2018 PMOC report, based on the 1 April 2018 schedule, forecast a Target RSD of 21 February 2022. This produced conditional Finish Slippage of 5.26% relative to the frozen target of 12 February 2021. The June 2018 PMOC report documented the April 2018 cost reassessment underlying the 9.39% EAC variance observed later in the primary October 2018 record. Later compatible reports showed the persistence of elevated cost and schedule conditions. An October 2019 target RSD of 14 March 2022 increased finish slippage to approximately 5.56%, but did not change the elevated classification. Therefore, the selection of the reporting period affected the earliest observable timing, but not the maximum risk band or the main cross-domain interpretation. No separate sensitivity series was constructed for Canarsie; its favorable-control interpretation is based on the selected project-level historical snapshots and the separately reported component-level evidence.

4.2. Case 1: NYCT ADA station improvements

The NYCT case showed a decline in the published completion forecast for the 149 St–Grand Concourse and Tremont Av accessibility improvement programme. In both July 2021 and January 2022, the committee records reflected the original substantial completion target of July 2023 and finish Slippage of 0.00% and a Low classification of 10 at both observations (MTA, 2021, 2022). The Independent Engineering Consultant review conducted in January 2022 had already identified a number of schedule blockers with station milestones, utility coordination and limitations on schedule float. These issues were documented without any corresponding change to the headline completion forecast. By 21 February 2023, the reported programme completion forecast had shifted to January 2024. Compared to the original 1 July 2023 baseline, this represented a delay of 184 days, equivalent to 17.29% of the baseline duration of 1,064 days. The unchanged IPCRF threshold classified the observation as critical (100), 130 days before the original completion target. The October 2024 report subsequently moved the forecast to March 2025, marking 609 days of slippage (57.24%) and keeping the Critical indicator active (MTA, 2023b, 2024).

Cost evidence was less continuous. The February 2023 report recorded both the contemporaneous project budget and EAC at $163 million, yielding an EAC variance of 0.00% on that reporting basis. As at October 2024, the budget is still $163 million, and the EAC is reported at $168 million, with a variance of 3.07% and a 'Watch' classification of 30. These observations could not support a continuous cost trend calculation (MTA, 2023b, 2024).

The case therefore demonstrates retrospective classification of disclosed schedule deterioration before the original completion benchmark, but not detection before the completion forecast itself changed. A verified project-wide actual completion date was unavailable in the audited evidence.

4.3. Case 2: Dulles Corridor Metrorail Phase 2

Dulles demonstrated divergence between the reported project-level cost forecast and other delivery-control conditions during the selected 2019 observations. The reported project budget and EAC remained approximately $2,778.236 million ($2.778 billion) in March, June, and the December-labelled report, producing an EAC variance of 0.00% and a Low classification of 10. Cost performance indices were 1.037, 1.042 and 1.025, respectively, also mapping to Low. These findings indicate stability in the reported project-level cost forecast, not an absence of underlying delivery pressure (FTA, 2019a–c). The reporting-period sensitivity analysis identified High Total Change Exposure before the March primary observation and an earlier Elevated Pending Change Exposure observation, but these additional records did not increase the maximum classification bands reported in the primary analysis (Supplementary Table S5).

Schedule evidence showed a different condition. In June and the December-labelled 2019 report, the risk-adjusted revenue-service date was 16 July 2020, 107 days beyond the accepted 31 March 2020 baseline. Using the documented Package A construction-start proxy, Finish Slippage was 4.35%, corresponding to Elevated (60). The selected milestone reconstruction identified later forecast dates for all 11 milestones examined in June and all 13 examined in December. Because these observations did not establish the complete project-wide milestone population, they were

retained as selected-population diagnostics rather than formally scored Milestone Delay Rate indicators.

Total Change Exposure was 11.62%, 11.83% and 11.85% across the three primary observations, remaining in the High band. Pending Change Exposure increased from 3.00% in March to 3.03% in June before declining to 1.31% in the December-labelled report. Thus, cumulative change exposure remained high despite a later reduction in unresolved pending exposure (FTA, 2019a–c).

Contractor records documented one missed contractual substantial-completion milestone in March and June and two in December, corresponding to selected-population proportions of 1/1, 1/1 and 2/2. These observations were retained as diagnostic evidence of realised contractual failures rather than formal Contractor Performance scores. Revenue service commenced on 15 November 2022, 959 days after the accepted 31 March 2020 baseline (WMATA, 2022). This outcome provides a retrospective comparison endpoint; the selected 2019 classifications did not predict the eventual 959-day delay.

4.4. Case 3: East Side Access / Grand Central Madison

East Side Access experienced concurrent cost and schedule pressures, as well as documented access and interface constraints. The EAC in April 2018 was $11,133.3 million, which exceeded the frozen programme budget of $10,177.8 million in June 2014 by 9.39%, corresponding to Elevated (60). This comparison used the frozen June 2014 historical budget and should not be interpreted as a comparison with the contemporaneously approved budget at every subsequent observation. Because the same EAC was carried across the selected 2018–2019 records, the repeated 9.39% value was not interpreted as continued cost acceleration.

Schedule reconstruction showed an forecast internal revenue-service date of 21 February 2022 in the October 2018 report, which was 374 days after the fixed target date of 12 February 2021. The May and September 2019 records reported a forecast internal RSD of 14 February 2022. Using the documented construction-start proxy, these observations produced finish slippage values of 5.26%, 5.17%, and 5.17% respectively, all of which were classified as elevated (60).

Programme schedule contingency declined from an approximately 994-day historical basis to 295 days in October 2018 and 302 days in the later selected observations. These values correspond to approximately 70.32% and 69.62% depletion of the original programme buffer. They were retained as descriptive schedule-contingency measurements rather than formally scored CPM Float Erosion because programme contingency and CPM float were not demonstrated to be equivalent.

Constructability records documented access restrictions, inter-contract coordination difficulties and delayed handoffs. The September 2019 coordination-point reconstruction contained 28 selected activities, including 14 unresolved later forecasts, three completed-late activities and one unactualised boundary-date item. The narrow 14/28 measure represented 50.0% of the selected population, but the absence of a verified project-wide interface denominator prevented formal Interface Failure Rate scoring (FTA, 2019d).

Limited shuttle service commenced on 25 January 2023 and full service on 27 February 2023. Full service occurred 745 days after the internal target and 76 days after the 13 December 2022 late/replanned RSD benchmark (MTA, 2023a, 2023c). These comparisons demonstrate how the selected schedule reference materially affects the apparent magnitude of delay.

4.5. Case 4: Canarsie Tube / L Project Tunnel Rehabilitation

Canarsie provided the favorable-outcome comparator for the four-case analysis. At the 29 September 2019 snapshot, the published forecast placed tunnel rehabilitation completion in April 2020, approximately three months ahead of the source-stated planned position. Under the frozen Version 1.2 Finish Slippage formula, negative schedule variance was floored at zero, producing 0.00% and Low (10). The 25 March 2020 record continued to anticipate completion in April 2020 and again supported Low (10). The tunnel work was subsequently completed on 26 April 2020, and later MTA documentation characterised the outcome as three months ahead of schedule.

The MTA 2Q2020 Traffic Light Report also provided eight component-level contingency indices that were formula-compatible with the IPCRF Contingency Stress Ratio when third-party contract completion was used as the available progress measure. Seven components mapped to Low and one, the Pump Room at 0.94, mapped to Watch; none reached Elevated, High or Critical. The same source reported component Cost Index values between 0.49 and 1.00 and zero months of quarter-over-quarter schedule variance for all eight components. These measures were retained as contextual evidence because their definitions were not identical to IPCRF EAC Variance or Finish Slippage.

Later MTA material reports that the tunnel rehabilitation was completed $100 million under budget. This favorable cost outcome was not used to construct an exact project-level EAC Variance because the public records contain different contract and programme scope bases. Risk, change, contractor-performance and constructability inputs were also insufficient for full scoring. Case 4 therefore provides an initial control comparison in which the reconstructable signals remained Low or modest, but one favorable case does not establish specificity or a false-positive rate.

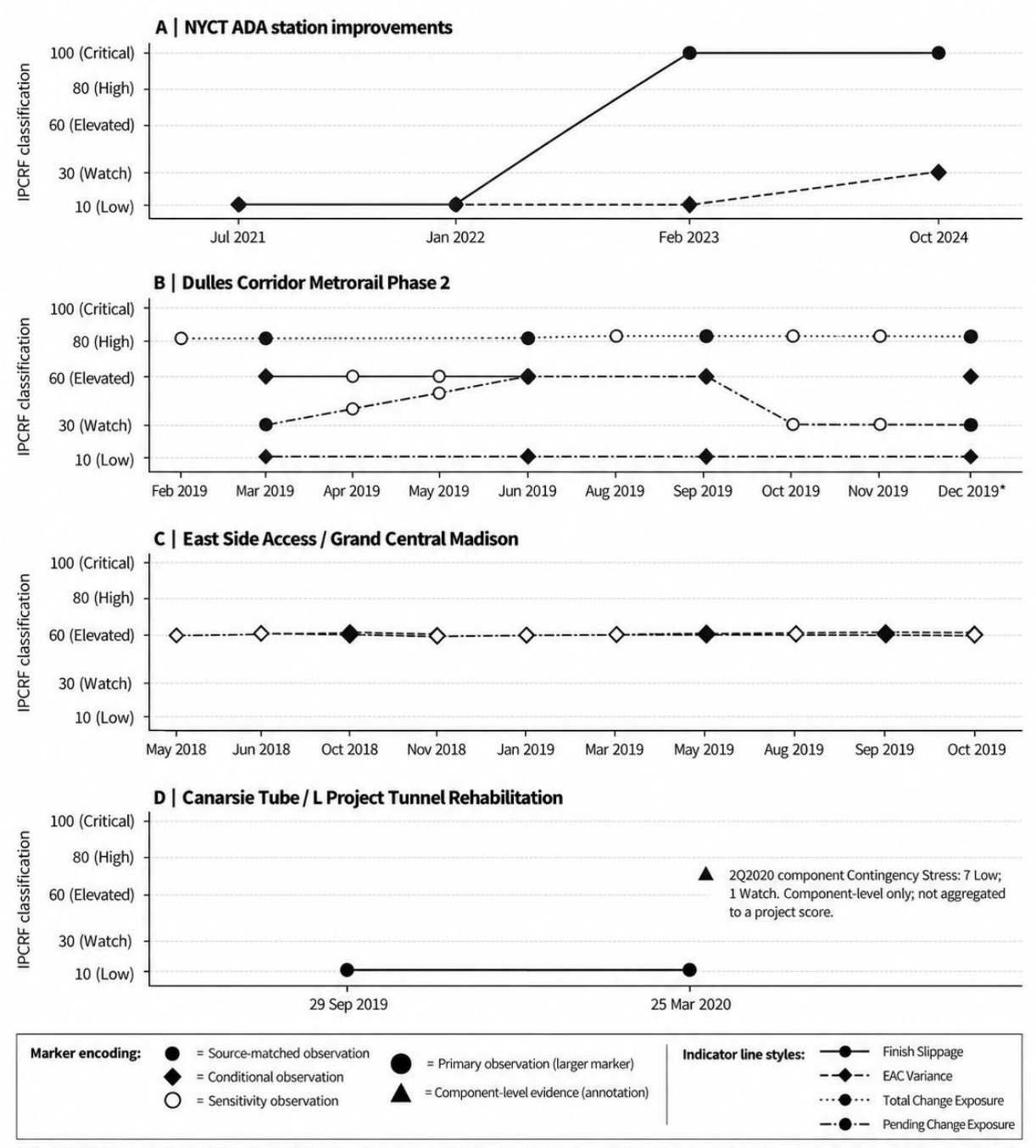


Figure 2. Primary and sensitivity trajectories of selected reconstructable IPCRF indicators across four historical cases

Note. IPCRF = Integrated Project Controls and Risk Forecasting Framework; EAC = estimate at completion. Values represent selected source-matched, conditional, or sensitivity indicator-level classifications rather than formal domain or composite scores. Filled markers denote primary observations, open markers denote sensitivity observations, and diamonds denote conditional observations. Connecting lines show chronology only and do not imply continuous measurement. For the Canarsie Tube/L Project, the two plotted Finish Slippage observations are project-level historical classifications, while the 2Q2020 Contingency Stress observations are component-level evidence and were not aggregated into a project-level score. No sensitivity series was constructed for Case Study 4. The December-labelled Dulles report contains relevant data through November 2019.

4.5. Cross-case warning timing and indicator divergence

Cross-case comparison revealed differences in the timing, severity and observability of project-control conditions. The reporting-period sensitivity analysis showed that some qualifying classifications were observable in additional monthly records before the originally selected primary observations. Accordingly, Table 3 reports the earliest compatible qualifying observations identified across the primary and sensitivity records rather than treating the first selected snapshot as the true onset of deterioration. These dates nevertheless remain earliest observed historical evidence within the reviewed records and should not be interpreted as definitive first-warning or first-public-disclosure dates.

Table 3: Historical indicator classifications, diagnostic observations and outcome benchmarks across the four U.S. transit capital projects

Project	Indicator or diagnostic observation	Earliest compatible observation	Classification/evidence status	Relevant benchmark/outcome	Temporal interpretation
NYCT ADA improvements	Finish Slippage: 17.29%	February 2023	Critical (100)	Original completion target: 1 July 2023	Observed 130 days before the original target, but after the completion forecast had already changed.
Dulles Phase 2	Total Change Exposure: 11.57%	February 2019 sensitivity record	High (80)	Accepted RSD: 31 March 2020; actual opening: 15 November 2022	Earlier than the March primary observation. Historical onset before February remains unestablished.
Dulles Phase 2	Finish Slippage: 6.02%	5 March 2019 schedule status reported in the February PMOC record	Elevated (60); conditional	Accepted RSD: 31 March 2020	148-day forecast difference normalised using the Package A construction-start proxy; not an accepted reconciled

					master schedule.
Dulles Phase 2	Pending Change Exposure: approximately 3.04%	April 2019 sensitivity record	Elevated (60)	Accepted RSD: 31 March 2020	Earlier Elevated crossing than the June primary record; later compatible observations returned to Watch.
Dulles Phase 2	Missed contractual substantial-completion milestone: 1/1	March 2019 primary record	Selected-population diagnostic; no formal score	Actual revenue-service commencement: 15 November 2022	Contractual failure had already occurred; evidence of contractor stress rather than advance detection of that failure.
Dulles Phase 2	Selected milestone forecasts later than reference dates	June 2019	Selected-population diagnostic; no formal score	Accepted RSD: 31 March 2020	Documents deterioration in selected milestones; does not establish a project-wide Milestone Delay Rate.
East Side Access	Finish Slippage: 5.26%	May 2018 report; 1 April 2018 schedule data	Elevated (60); conditional	Internal RSD target: 12 February 2021	Earlier than the October 2018 primary observation; classification depends on the construction-start proxy.
East Side Access	EAC variance against frozen 2014 budget: 9.39%	June 2018 report documenting April 2018 reassessment	Elevated (60); conditional	Full service: 27 February 2023	Earlier than the October 2018 primary observation; classification depends on the frozen

					historical cost comparator.
East Side Access	Unresolved later forecasts among selected coordination points: 14/28	September 2019	Selected-population diagnostic; no formal score	Full service: 27 February 2023	Selected coordination-point population only; complete project-wide interface denominator unavailable.
Canarsie Tube / L Project	Finish Slippage: 0.00%	29 September 2019	Low (10)	Tunnel work completed 26 April 2020; later MTA material reports three months ahead of schedule	Remained Low again at the 25 March 2020 snapshot. No Elevated, High or Critical schedule classification was observed.
Canarsie Tube / L Project	Component Contingency Stress: 0.00–0.94 across eight components	2Q2020	7 Low; 1 Watch; component-level mapping	Later official outcome reported $100 million under budget	No component reached Elevated, High or Critical. Values are not aggregated into a project-level Cost score.

Note. IPCRF = Integrated Project Controls and Risk Forecasting Framework; EAC = estimate at completion; RSD = revenue-service date. "Earliest compatible qualifying observation" refers to the earliest qualifying classification identified across the primary dataset and reporting-period sensitivity review. It does not necessarily represent the true historical onset of deterioration or first public disclosure. Conditional classifications depend on stated baseline, scope or normalisation assumptions. Selected-population diagnostics are not formal IPCRF risk scores. Temporal intervals are descriptive differences between specified historical information dates and relevant benchmarks or outcomes and should not be interpreted as validated public-warning lead times. Supplementary Table S5 documents the reporting-period sensitivity assessment. Sources: FTA historical PMOC records, MTA project records, WMATA (2022), and Supplementary Tables S1–S5.

The most distinct Low to Critical Finish Slippage occurred on records provided by NYCT. Dulles showed simultaneous financial stability and non-cost delivery stress, while East Side Access demonstrated that internal schedule deterioration could coexist with a comparatively smaller miss against the public service commitment.

The Dulles comparison was very useful, as variance on the project level was Low, cumulative change exposure was High, and contractual milestones had already been missed. East Side access also indicated a fairly consistent carried-forward EAC forecast, as well as access and coordination challenges. These conditions are consistent with what was documented in the original oversight records, though, and not seen as "abnormal" delivery issues.

Case and observation		Finish slippage	EAC variance	Cost performance index	Contingency stress	Total change exposure	Pending change exposure
NYCT	Jul 2021	10 (L)	10† (L)	—	—	—	—
	Jan 2022	10 (L)	10† (L)	—	—	—	—
	Feb 2023	100 (C)	10 (L)	—	—	—	—
	Oct 2024	100 (C)	30 (W)	—	—	—	—
Dulles	Mar 2019	—	10 (L)	10 (L)	10† (L)	80 (H)	30 (W)
	Jun 2019	60† (E)	10 (L)	10 (L)	10† (L)	80 (H)	60 (E)
	Dec 2019*	60† (E)	10 (L)	10 (L)	10† (L)	80 (H)	30 (W)
ESA	Oct 2018	60† (E)	60† (E)	—	—	—	—
	May 2019	60† (E)	60† (E)	—	—	—	—
	Sep 2019	60† (E)	60† (E)	—	—	—	—
Canarsie	Sep 2019	10 (L)	—	—	—	—	—
	Mar 2020	10 (L)	—	—	—	—	—

Classification key (value and label)

10 (L) Low | 30 (W) Watch | 60 (E) Elevated | 80 (H) High | 100 (C) Critical

Symbol key

† = conditional measurement involving baseline, scope or normalisation restrictions

— = unavailable or withheld project-level classification

Figure 3. Primary project-level indicator classifications and evidence availability across four cases

Note. IPCRF = Integrated Project Controls and Risk Forecasting Framework; EAC = estimate at completion; ESA = East Side Access. Values represent selected source-matched or conditional indicator-level classifications rather than formal domain or composite scores. The dagger (†) identifies conditional measurements involving baseline, scope, or normalisation restrictions. A dash (—) indicates an unavailable or withheld project-level indicator classification and must not be interpreted as Low risk. Selected contractual failures, milestone-delay proportions, programme-contingency depletion, and coordination-point diagnostics are excluded from formal risk classification. Case Study 4's eight 2Q2020 Contingency Stress mappings are component-level observations and are therefore excluded from this project-level matrix; they are reported separately in Table 3.

4.6. Data completeness, rebase lining and source-related findings

The historical reconstruction identified three common problems with formal scoring: exclusion of quantitative fields, not all baseline definitions were consistent, and there was no consistency in the dates associated with the reports and the information available. There was an especially significant shortfall in probabilities within the risk register, complete mitigation histories, as well as contractor compliance denominators and project-wide constructability records. Some domains provided

diagnostic calculations, and the required fields completeness threshold could not be reliably identified for cases of all types. For this reason, a full PDRS, as well as a full FCS, were not reported (Yusuf, 2026).

The interpretation was affected by the selection of the baseline. Dulles achieved the accepted revenue-service baseline figure in existing budgets for March 2020, in addition to its earlier forecast and later risk adjustment. The internal and public benchmarks of East Side Access couldn't be merged, and the historical and new cost bases of NYCT also could not be merged. Some Dulles reports were finalised after their nominal reporting periods, and the original public release dates remained unverified. These limitations did not allow for the identification of definite public warning lead times. Table 4 synthesises the evidence addressing the four research questions across the four cases, including the Canarsie Tube/L Project as a favorable-outcome comparator, while preserving the principal evidential boundaries affecting cross-project interpretation.

Table 4. Cross-case synthesis of empirical findings and evidential boundaries across four U.S. transit capital projects

Research question	NYCT ADA station improvements	Dulles Phase 2	East Side Access / Grand Central Madison	Canarsie Tube / L Project Tunnel Rehabilitation	Cross-case interpretation
RQ1. Historical characterisation of project-control conditions	Critical Finish Slippage was observed before the original completion target, but only after the published completion forecast had already changed.	Multiple indicators documented delivery stress before revenue-service commencement, including High Total Change Exposure, conditional Elevated Finish Slippage and already-realised contractual milestone failures.	Elevated cost and conditional schedule classifications, together with documented access and interface constraints, were observed before full-service commencement.	Finish Slippage was Low at both project-level historical snapshots. Eight component-level Contingency Stress mappings comprised seven Low and one Watch; none was Elevated or above. Later official records document completion ahead of schedule and under budget.	The three deterioration-enriched cases and the favorable comparator produced materially different historical indicator patterns under the same frozen rules. This strengthens the comparative design but does not establish statistical discrimination.

RQ2. Timing of observed qualifying classifications	A Low-to-Critical Finish Slippage transition was observed within the reviewed records. The Critical classification occurred before the original completion benchmark, but after the published completion forecast had already changed.	High Total Change Exposure was identifiable in the February 2019 sensitivity record. Conditional Elevated Finish Slippage was also identified in earlier compatible records. Some contractor failures had already occurred and were therefore diagnostic rather than advance warnings.	Conditional Elevated Finish Slippage was identifiable in the May 2018 sensitivity record, while the April 2018 cost reassessment documented in June 2018 supported an earlier Elevated EAC Variance classification.	No Elevated, High or Critical classification was observed in the two project-level schedule snapshots or the eight component contingency mappings. An adverse warning interval is therefore not calculated for this favorable comparator.	Severe qualifying classifications were observed only where supported by the reviewed records; their absence in the favorable comparator provides an initial check against mechanical over-triggering, not an estimate of specificity.
RQ3. Contribution of multi-domain information	Limited quantitative coverage prevented a formal integrated assessment; the strongest reconstructable evidence was schedule-related, with limited cost observations.	Low project-level EAC variance coexisted with High Total Change Exposure, conditional schedule deterioration and documented contractor failures, demonstrating divergence between the headline cost forecast and other project-control dimensions.	Cost and schedule pressure occurred alongside documented access and coordination constraints, providing information beyond the selected completion-date benchmarks.	Favorable Finish Slippage and component contingency classifications were directionally consistent with MTA cost indices at or below 1.00 and zero months of quarter-over-quarter component schedule variance. The latter measures remain contextual because their definitions differ from	Cross-domain interpretation distinguishes both hidden delivery stress, as in Dulles, and comparatively favorable control conditions, as in Canarsie. Four cases remain insufficient to establish a general discriminatory signature.

				formal IPCRF indicators.	
RQ4. Effect of data completeness, baseline management and reporting structure	Historical cost bases were not fully continuous, and several IPCRF domains lacked reconstructable quantitative inputs.	Missing float, quantitative risk and contractor-compliance fields, together with multiple schedule references and differing report/data dates, restricted formal scoring.	Distinct historical cost and schedule benchmarks, incomplete interface denominators and proxy-dependent schedule measures constrained formal scoring and direct comparison.	Project-level BAC/EAC history and quantitative Risk, Change, Contractor Performance and Constructability inputs were incomplete. Component- and scope-specific values were retained separately; no full domain, PDRS or FCS was inferred.	Reporting density, baseline governance, field completeness and measurement compatibility remain major constraints on direct cross-project comparison. The favorable comparator improves outcome contrast but does not remove these data limitations or permit unsupported domain, PDRS or FCS aggregation.

Note. IPCRF = Integrated Project Controls and Risk Forecasting Framework; BAC = budget at completion; EAC = estimate at completion; PDRS = Project Delivery Risk Score; FCS = Forecast Confidence Score. Findings refer to the selected historical records and compatible sensitivity observations. "Earliest observed qualifying classification" refers to the earliest supported classification identified within the reviewed records and does not necessarily represent the true onset of deterioration or first public disclosure. Conditional classifications depend on stated baseline, scope or normalisation assumptions, while selected-population diagnostics are not formal IPCRF scores. Case 4 serves as a favorable-outcome comparator and was analysed using the same frozen IPCRF Version 1.2 rules applied to Cases 1–3. Its eight 2Q2020 Contingency Stress mappings are component-level observations and were not aggregated into a project-level Cost score. One favorable comparator does not establish specificity, false-positive performance or predictive discrimination. All four cases represent retrospective historical applications rather than prospective predictive validation.

5. Discussion

5.1. Interpretation of historical warning behaviour

The historical findings show contrasting indicator behaviour across the four cases. NYCT, Dulles and East Side Access contained different forms of reconstructable delivery stress, whereas the Canarsie Tube/L Project remained Low on Finish Slippage at both project-level historical snapshots and showed seven Low and one Watch component-level Contingency Stress observation. No Canarsie classification reached Elevated, High or Critical. The favorable comparator therefore provides an initial control check against the possibility that the frozen framework mechanically produces severe classifications whenever it is applied retrospectively to

a major transit project. This remains exploratory evidence: one favorable case cannot establish specificity, false-positive performance or predictive discrimination.

Batselier and Vanhoucke (2015) demonstrated that project-duration forecasting methods differ in terms of their accuracy and timeliness when assessed using real project data. For this reason, later research by the same researchers (Batselier & Vanhoucke, 2017) demonstrated that adding recent performance data and reference-class forecasting can enhance the cost and time approximation.

Likewise, de Andrade et al. (2019) highlighted the significance of scheduling adherence in the comparison of the methods to predict the duration of projects. The current findings help to further call attention to the interpretation of historical warning classifications, such that a signal can be accurate even if deterioration precedes the signal's disclosure.

The frequency of reporting also influences the 'warning interval' observed. More continuous tracking has been found, as conventional periodic measurements of earned value may underestimate performance changes between reporting points (Ngo et al., 2022). As a result, the observations used in this study are the first observed warnings for qualification and not the historical first warnings. Furthermore, the presence of a warning does not indicate whether an intervention would have prevented deterioration. Haji-Kazemi et al. (2015) identified organisational barriers that can prevent project managers from responding effectively to recognised warning signs. Thus, IPCRF's historic classifications should not be interpreted as proof of efficacy or predictive power for interventions but rather instead as a reflection of an observed mode of delivery.

5.2. Contribution of integrated project-control information

The differences between the headline measure of cost and the conditions of the contractor, schedule and interfaces show how much information can be extracted by studying project control domains together. Dulles provided the clearest example: project-level EAC variance remained low, while cumulative change exposure was high, and selected contractual milestones had already been missed. While these conditions were documented in the original oversight reports, their simultaneous classification demonstrated how distinct control measures could be compared within a common analytical structure. Love et al. (2017) looked at Australian rail infrastructure projects and found scope changes are a significant driver of cost growth. This shows the need, again, to track change exposure along with expenditure projections.

Canarsie provides the converse pattern. Its reconstructable schedule and component contingency signals remained favorable or modest, and the available contextual MTA cost and schedule measures were directionally consistent with the later favorable outcome. This contrast shows why integrated interpretation should examine both the presence and absence of cross-domain stress rather than assume deterioration as the expected result.

Likewise, Soliman et al. (2025) investigated the incorporation of risk management and earned-value information for identifying project warning conditions. The rationale for risk integration

and integration of earned-value information, as discussed in the Dulles findings, is consistent with this concept, but only partially supported because the earned-value information had no hard numbers to support the analysis of the fuller scope of the multi-domain mechanism of the IPCRF.

East Side Access also showed how important access and interface conditions are when interpreting schedule performance. Lagos et al. (2026) studied the contributions of collaboration indicators, in addition to the production planning and earned value indicators, to the prediction of schedule performance. Similarly, Nguyen (2020) found that construction planning and schedule control functions were significant in an empirical study of finished projects. The results justify the study of operational and managerial information as well as traditional operational measures.

However, Locatelli et al. (2017) found that project characteristics and stakeholder conditions were associated with performance in a study of 30 European transport infrastructure projects. Observed differences in performance must therefore be understood instead as being specific to a particular project's delivery and reporting process and not as indicating that the same warning patterns would be observed in cases of transit projects.

5.3. Practical implications for public-transit project monitoring

The results highlight the need for integrated baseline management in early warning system implementation, consistent baseline management, structured interface records, and comprehensive project control data. For transit agencies, this means that indicator calculations must be traceable to compatible source values, reporting periods and approved baselines. The Federal Transit Administration (FTA, 2025) underscores the importance of managing scope, schedule, cost and risk in a coordinated manner across the full life cycle of capital projects.

The present study confirms the need for continual historical measurement and documentation for this coordination. The projects of Dulles and East Side Access demonstrated that the choice of original, revised or public completion baseline can have a significant effect on what constitutes schedule deterioration. The U.S. Government Accountability Office (GAO, 2015) has defined the following practices as necessary for project delivery monitoring: baseline maintenance, reliable schedule logic, and progress updates.

Agencies applying IPCRF should retain original and approved revised baselines alongside the forecasts associated with each reporting period rather than overwrite earlier reference points. Cost information requires the same discipline: each EAC, actual-cost measure and change value should be linked to a documented budget and scope basis so that unlike scopes are not combined into artificial variances (GAO, 2020). Case 4 illustrates this requirement particularly clearly because its public records contain contract-, tunnel- and broader programme-level cost figures that cannot defensibly be treated as one cost series.

The second implication has to do with the structured contractor and interface data collection. Within the narrative accounts, there were challenges of various sorts throughout, but often lacked denominators, event dates, and status histories on which to calculate indicators.Project-control

systems should record identifiable milestones, change events, access restrictions, interface handoffs, ownership and resolution dates. According to De Marco & Narbaev (2021), the factors that influence cost and schedule performance in tunnel megaprojects are infrastructure complexity and contracting arrangements.

These technical requirements also depend on institutional coordination. Examining California High-Speed Rail, Sekasi et al. (2025) demonstrated how governance, contracting and organisational conditions interact in megaproject delivery. A prospective pilot for the IPCRF should therefore include standardised data gathering, along with specified responsibilities for warning review, source quality review and correction activities. Where the required completeness threshold is not met, composite scores should be withheld rather than inferred.

5.4. Study strengths, limitations and future testing

The study's main strength lies in its use of authentic historical project records. The conclusions are, however, constrained by the incompleteness of reporting periods and domain coverage, and by using the framework retroactively. An independent reconstruction of the available indicators kept the separation of documented inputs and derived classifications and subsequent outcomes. The four-case design now includes one favorable-outcome comparator in addition to the three deterioration-enriched cases, reducing the original outcome-selection imbalance. However, one comparator is insufficient to estimate specificity, false-positive performance or discriminatory accuracy, and broader testing across multiple stable, recovered and adversely performing projects remains necessary.

The primary reporting periods were selected and therefore did not independently establish the earliest historical onset of deterioration. A secondary review of compatible monthly PMOC records identified earlier qualifying observations for several Dulles and East Side Access indicators, confirming that the selected snapshots could understate onset timing. However, the additional records did not produce higher classification bands or reverse the principal cross-domain patterns. The sensitivity analysis therefore supports the robustness of the severity interpretation while reinforcing the distinction between earliest observed evidence and true warning onset. Original public-release dates could not be consistently verified, and missing risk, contractor and interface fields prevented complete framework scoring. Some schedule and constructability measurements also depended on proxies whose equivalence to formal IPCRF indicators could not be established. Although Version 1.2 was fixed before the present case-study extraction and scoring, the historical projects and their principal outcomes predated the framework; the study therefore remains a retrospective historical application rather than prospective validation.

Bayram and Al-Jibouri (2016) showed the application of reference-class forecasting to historical construction-cost outcomes. This illustrates the importance of using comparative project evidence to evaluate forecasting approaches.

The GAO (2014) also documented significant variations in transit capital project development periods and evolving project estimates, emphasising the importance of appropriate baseline and project context comparisons. Future validation should therefore include consecutive reporting periods, verified source availability dates, complete field-level data, and projects with different delivery outcomes. An assessment of warning response before and after onset, persistence and the response to management of the warning should be undertaken by a prospectively specified pilot before deriving thresholds or claiming predictive performance.

6. Conclusion

Historical back-testing across four U.S. transit capital projects showed that selected IPCRF Version 1.2 indicators can reconstruct contrasting project-control conditions without forcing all cases toward severe risk. NYCT, Dulles and East Side Access exhibited different combinations of schedule, change, contractor and interface pressure, whereas Canarsie remained Low on project-level Finish Slippage and showed seven Low and one Watch component-level Contingency Stress classification across its reconstructed evidence. This favorable comparator improves the outcome balance of the study but does not establish specificity, false-positive rates, predictive validity or operational effectiveness. Incomplete fields, incompatible cost scopes, changing baselines and proxy-dependent measurements continued to prevent formal six-domain PDRS and FCS assessment. The evidence therefore supports retrospective indicator-level historical reconstruction across contrasting delivery conditions. Future work should use prospectively defined outcomes, consecutive time-stamped observations and a larger sample of adverse, stable and recovered projects before predictive performance is inferred.

7. Data Availability and Reproducibility

The frozen IPCRF Version 1.2 specification is publicly archived on Zenodo at DOI 10.5281/zenodo.23028918**.**

The reproducibility packages for Case Studies 1–4 are archived separately at DOI 10.5281/zenodo.23088088**,** 10.5281/zenodo.23088683**,** 10.5281/zenodo.23088858**,** and 10.5281/zenodo.23088970, respectively.

The complete four-case evidence collection is archived at DOI 10.5281/zenodo.23089716**.**

The Case 4 archive includes the technical report, reproducibility workbook/raw data, Data.csv and Source_Provenance.csv used for the historical reconstruction. All cases were analysed using the frozen Version 1.2 specification; no later scoring-rule clarification was retrospectively applied to the reported results.

8. Declarations

Data Availability

The data and materials supporting this study are publicly available. The frozen IPCRF Version 1.2 specification is available at https://doi.org/10.5281/zenodo.23028918. The reproducibility packages for Case Studies 1 through 4 are available at https://doi.org/10.5281/zenodo.23088088, https://doi.org/10.5281/zenodo.23088683, https://doi.org/10.5281/zenodo.23088858, and https://doi.org/10.5281/zenodo.23088970, respectively. The complete four case evidence collection is available at https://doi.org/10.5281/zenodo.23089716. The Case Study 4 archive includes the technical report, reproducibility workbook and raw data, Data.csv, and Source_Provenance.csv used for the historical reconstruction. All cases were analyzed using the frozen IPCRF Version 1.2 specification, and no later scoring rule clarification was retrospectively applied to the reported results.

Funding

This review was funded by the authors.

Conflict of Interest

All the authors have no competing interests.

Clinical Trial Number

Not applicable

Ethical Approval

Not applicable. The manuscript is purely a scoping review of scholarly literature and does not involve any physical or virtual contact with any human for data collection.

Consent to participate

Not applicable

Consent to publish

Not applicable

Author Contribution

M.Y. conceptualized the study, developed the IPCRF framework, designed the historical back-testing methodology, collected and curated the public-source project data, conducted the four case reconstructions and analyses, interpreted the findings, prepared the manuscript and supporting tables and figures, and revised the manuscript. M.Y. reviewed and approved the final manuscript and accepts responsibility for the integrity of the work

References

Batselier, J., & Vanhoucke, M. (2015). Evaluation of deterministic state-of-the-art forecasting approaches for project duration based on earned value management. International Journal of Project Management, 33(7), 1588–1596. https://doi.org/10.1016/j.ijproman.2015.04.003

Batselier, J., & Vanhoucke, M. (2017). Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. International Journal of Project Management, 35(1), 28–43. https://doi.org/10.1016/j.ijproman.2016.10.003

Bayram, S., & Al-Jibouri, S. (2016). Application of reference class forecasting in Turkish public construction projects: Contractor perspective. Journal of Management in Engineering, 32(3), Article 05016002. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000421

Cho, K., Ahn, S., Park, K., & Kim, T. W. (2021). Schedule delay leading indicators in precast concrete construction projects: Qualitative comparative analysis of Korean cases. Journal of Management in Engineering, 37(4), Article 04021024. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000915

de Andrade, P. A., Martens, A., & Vanhoucke, M. (2019). Using real project schedule data to compare earned schedule and earned duration management project time forecasting capabilities. Automation in Construction, 99, 68–78. https://doi.org/10.1016/j.autcon.2018.11.030

De Marco, A., & Narbaev, T. (2021). Factors of schedule and cost performance of tunnel construction megaprojects. The Open Civil Engineering Journal, 15, 38–49. https://doi.org/10.2174/1874149502115010038

Federal Transit Administration. (2014). PMOC comprehensive monthly report: East Side Access (MTACC–ESA) project, December 2014 [Project Management Oversight Contractor report]. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/ESA_Monthly_PMOC_Report_-_December_2014.pdf

Federal Transit Administration. (2018). PMOC monthly report: East Side Access (MTACC–ESA) project, October 2018 [Project Management Oversight Contractor report]. https://www.transit.dot.gov/sites/fta.dot.gov/files/2021-07/esa-monthly-pmoc-report-October-2018.pdf

Federal Transit Administration. (2019a). Dulles Corridor Metrorail Project Phase 2: Wiehle Avenue Station to Ashburn Station, December 2019 [Project Management Oversight Contractor monthly monitoring report]. https://www.transit.dot.gov/foia/dulles-corridor-metrorail-project-phase-2-wiehle-avenue-station-ashburn-station-december-2019

Federal Transit Administration. (2019b). Dulles Corridor Metrorail Project Phase 2: Wiehle Avenue Station to Ashburn Station, June 2019 [Project Management Oversight Contractor monthly

monitoring report]. https://www.transit.dot.gov/foia/dulles-corridor-metrorail-project-phase-2-wiehle-avenue-station-ashburn-station-june-2019

Federal Transit Administration. (2019c). Dulles Corridor Metrorail Project Phase 2: Wiehle Avenue Station to Ashburn Station, March 2019 [Project Management Oversight Contractor monthly monitoring report]. https://www.transit.dot.gov/foia/dulles-corridor-metrorail-project-phase-2-wiehle-avenue-station-ashburn-station-march-2019

Federal Transit Administration. (2019d). PMOC comprehensive monthly report: East Side Access (MTACC–ESA) project, September 2019 [Project Management Oversight Contractor report]. https://www.transit.dot.gov/sites/fta.dot.gov/files/2021-07/esa-monthly-pmoc-report-september-2019.pdf

Federal Transit Administration. (2019e). PMOC monthly report: East Side Access (MTACC–ESA) project, May 2019 [Project Management Oversight Contractor report]. https://www.transit.dot.gov/sites/fta.dot.gov/files/2021-07/esa-monthly-pmoc-report-May-2019.pdf

Federal Transit Administration. (2024, December 4). Project management oversight procedures. U.S. Department of Transportation. https://www.transit.dot.gov/regulations-and-guidance/project-management-oversight-procedures

Federal Transit Administration. (2025). Project and construction management guidelines (January 2025 ed.). U.S. Department of Transportation. https://www.transit.dot.gov/sites/fta.dot.gov/files/2025-02/Project-and-Construction-Management-Guidelines-January-2025.pdf

Gao, N., & Touran, A. (2020). Cost overruns and formal risk assessment program in US rail transit projects. Journal of Construction Engineering and Management, 146(5), Article 05020004. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001827

González-Cruz, M.-C., Ballesteros-Pérez, P., Lucko, G., & Zhang, J.-X. (2022). Critical duration index: Anticipating project delays from deterministic schedule information. Journal of Construction Engineering and Management, 148(11), Article 04022116. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002387

Haji-Kazemi, S., Andersen, B., & Klakegg, O. J. (2015). Barriers against effective responses to early warning signs in projects. International Journal of Project Management, 33(5), 1068–1083. https://doi.org/10.1016/j.ijproman.2015.01.002

Lagos, C. I., Herrera, R. F., Mac Cawley, A. F., & Alarcón, L. F. (2024). Predicting construction schedule performance with last planner system and machine learning. Automation in Construction, 167, Article 105716. https://doi.org/10.1016/j.autcon.2024.105716

Lagos, C. I., Herrera, R. F., Mac Cawley, A., & Alarcón, L. F. (2026). Prediction of the project schedule performance outcome with last planner system and social network analysis indicators. Engineering, Construction and Architectural Management, 33(2), 1116–1141. https://doi.org/10.1108/ECAM-03-2025-0424

Locatelli, G., Invernizzi, D. C., & Brookes, N. J. (2017). Project characteristics and performance in Europe: An empirical analysis for large transport infrastructure projects. Transportation Research Part A: Policy and Practice, 98, 108–122. https://doi.org/10.1016/j.tra.2017.01.024

Love, P. E. D., Zhou, J., Edwards, D. J., Irani, Z., & Sing, C. P. (2017). Off the rails: The cost performance of infrastructure rail projects. Transportation Research Part A: Policy and Practice, 99, 14–29. https://doi.org/10.1016/j.tra.2017.02.008

Metropolitan Transportation Authority. (2021). Capital Program Oversight Committee meeting: July 19, 2021 [Committee meeting materials]. https://www.mta.info/document/44206

Metropolitan Transportation Authority. (2022). Capital Program Committee meeting: January 24, 2022 [Committee meeting materials]. https://www.mta.info/document/72541

Metropolitan Transportation Authority. (2023a). Annual comprehensive financial report for the years ended December 31, 2022 and 2021. https://www.mta.info/document/114381

Metropolitan Transportation Authority. (2023b). Capital Program Committee meeting: February 21, 2023 [Committee meeting materials]. https://www.mta.info/document/105871

Metropolitan Transportation Authority. (2023c, June 21). MTA commuter railroads heading into summer with record total of 460,000 riders in single day [Press release]. https://www.mta.info/press-release/mta-commuter-railroads-heading-summer-record-total-of-460000-riders-single-day

Metropolitan Transportation Authority. (2024). Capital Program Committee meeting: October 28, 2024 [Committee meeting materials]. https://www.mta.info/document/155071

Muruganandan, K., Davies, A., Denicol, J., & Whyte, J. (2022). The dynamics of systems integration: Balancing stability and change on London's Crossrail project. International Journal of Project Management, 40(6), 608–623. https://doi.org/10.1016/j.ijproman.2022.03.007

Ngo, K. A., Lucko, G., & Ballesteros-Pérez, P. (2022). Continuous earned value management with singularity functions for comprehensive project performance tracking and forecasting. Automation in Construction, 143, Article 104583. https://doi.org/10.1016/j.autcon.2022.104583

Nguyen, L. H. (2020). Empirical analysis of a management function's failures in construction project delay. Journal of Open Innovation: Technology, Market, and Complexity, 6(2), Article 26. https://doi.org/10.3390/joitmc6020026

Nguyen, T., Chileshe, N., Ho, D. T., Nguyen, V. T., & Tran, Q. P. (2024). Significant risks that trigger cost overruns and delays in urban rail projects: A typical case study of Vietnam. Built Environment Project and Asset Management, 14(2), 278–295. https://doi.org/10.1108/BEPAM-01-2023-0027

Sekasi, J., Harrod, S., & Andrade, V. (2025). Governance, contracts, and culture: A triadic approach to megaproject performance. Project Leadership and Society, 6, Article 100197. https://doi.org/10.1016/j.plas.2025.100197

Shafaat, A., Marbouti, F., & Mahfouz, T. (2023). Early warning system for highway construction projects using GA-SVM. International Journal of Construction Management, 23(14), 2348–2357. https://doi.org/10.1080/15623599.2022.2056805

Shrestha, B., Shrestha, P. P., Maharjan, R., & Gransberg, D. (2022). Cost, change order, and schedule performance of highway projects. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 14(1), Article 04521044. https://doi.org/10.1061/(ASCE)LA.1943-4170.0000523

Soliman, E., Alrasheed, K. A., Alghanim, S., & Morsi, E. (2025). Integrating risk management and earned value framework to detect early warning signs – Case study. Journal of Engineering Research, 13(3), 2109–2116. https://doi.org/10.1016/j.jer.2024.05.029

U.S. Government Accountability Office. (2014). Public transit: Length of development process, cost estimates, and ridership forecasts for capital-investment grant projects (GAO-14-472). https://www.gao.gov/products/gao-14-472

U.S. Government Accountability Office. (2015). GAO schedule assessment guide: Best practices for project schedules (GAO-16-89G). https://www.gao.gov/products/gao-16-89g

U.S. Government Accountability Office. (2020). Cost estimating and assessment guide: Best practices for developing and managing program costs (GAO-20-195G). https://www.gao.gov/products/gao-20-195g

U.S. Government Accountability Office. (2023). Capital investment grants program: Cost predictions have improved, but the pandemic complicates assessing ridership predictions (GAO-23-105479). https://www.gao.gov/products/gao-23-105479

Washington Metropolitan Area Transit Authority. (2022, November 15). Metro launches Silver Line Extension with opening ceremony; welcomes customers to #RideSilver to six new stations [Press release]. https://www.wmata.com/news/silver-line-grand-opening.html

Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE Publications.

Mohammed Adeyemi Yusuf. (2026). Integrated Project Controls and Risk Forecasting Framework (IPCRF) for State-of-Good-Repair Capital Projects in U.S. Passenger Rail and Public Transit. Version 1.2. Zenodo. https://doi.org/10.5281/zenodo.23028918

Zhang, S., & Wang, X. (2023). Quantifying schedule delay risk in construction projects: A data-driven approach with BIM and probabilistic reliability analysis. Advances in Civil Engineering, 2023, Article 5525655. https://doi.org/10.1155/2023/5525655

Supplementary Table S1. Reconciled historical IPCRF indicator calculations, classifications and evidential boundaries across four U.S. transit capital projects

The expanded evidence register contains 75 records: 20 source-matched project-level indicators, 13 conditional project-level indicators, seven selected-population diagnostics, eight project-level descriptive measurements, one withdrawn calculation, eight Case 4 component-level mapped indicators, 16 Case 4 component-level contextual metrics and two Case 4 outcome-state checks. Thirty-three project-level indicator classifications were retained for graphical comparison. Component-level and outcome-state records are reported separately and are not included in the 276 potential project-level indicator-period cells.

Panel A. Dulles Corridor Metrorail Phase 2

Reporting period	Domain and indicator	Historical inputs or calculation basis	Recalculated value	IPCRF classification	Evidence status and interpretation
March 2019	Cost: EAC Variance	EAC and budget: $2,778.236 million each	0.00%	Low (10)	Source-matched indicator
March 2019	Cost: CPI	EV: $2,137.440m; actual cost: $2,060.324m	1.03743	Low (10)	Source-matched indicator
March 2019	Cost: Contingency Stress	Original contingency: $551.451m; remaining: $260.705m; physical progress: 86%	0.61307	Low (10), conditional	Potential mismatch between programme contingency and physical-construction progress bases
March 2019	Change: Total Change Exposure	Approved: $239.459m; pending: $83.255m; budget: $2,778.236m	11.62%	High (80)	Source-matched indicator
March 2019	Change: Pending Change Exposure	Pending changes: $83.255m; budget: $2,778.236m	3.00%	Watch (30)	Source-matched indicator
March 2019	Contractor: Missed Milestones	One missed milestone among one selected contractual milestone	1/1 (100%)	Not formally scored	Selected-population diagnostic; contractual failure already realised
June 2019	Cost: EAC Variance	EAC and budget: $2,778.236m each	0.00%	Low (10)	Source-matched indicator
June 2019	Cost: CPI	EV: $2,212.175m; actual cost: $2,123.289m	1.04186	Low (10)	Source-matched indicator

June 2019	Cost: Contingency Stress	Original contingency: $551.451m; remaining: $256.718m; physical progress: 89%	0.60053	Low (10), conditional	Scope compatibility not independently certified
June 2019	Schedule: Finish Slippage	Risk-adjusted RSD: 16 July 2020; accepted baseline: 31 March 2020; duration proxy: 2,458 days	4.35%	Elevated (60), conditional	Package A construction-start proxy; forecast was not an accepted reconciled master schedule
June 2019	Schedule: Milestone Delay Rate	Eleven selected milestones with later forecast dates	11/11 (100%)	Not formally scored	Selected milestones only; complete project-wide denominator unavailable
June 2019	Change: Total Change Exposure	Approved: $244.229m; pending: $84.319m	11.83%	High (80)	Source-matched indicator
June 2019	Change: Pending Change Exposure	Pending changes: $84.319m	3.03%	Elevated (60)	Source-matched indicator
June 2019	Change: Change Trend Growth	Change exposure compared with March 2019	1.81%	Not formally scored	Descriptive period-to-period change; interval normalisation not established
June 2019	Contractor: Missed Milestones	One missed milestone among one selected contractual milestone	1/1 (100%)	Not formally scored	Selected-population diagnostic; failure already realised
December 2019	Cost: EAC Variance	EAC and budget: $2,778.236m each	0.00%	Low (10)	Source-matched indicator; relevant financial information through November 2019
December 2019	Cost: CPI	EV: $2,308.593m; actual cost: $2,251.603m	1.02531	Low (10)	Source-matched indicator

December 2019	Cost: Contingency Stress	Original contingency: $551.451m; remaining: $236.592m; physical progress: 93%	0.61394	Low (10), conditional	Scope compatibility not independently certified
December 2019	Cost: Cost Trend Acceleration	Three selected, non-consecutive reports	0.00	Withdrawn	Does not satisfy a verified consecutive monthly trend requirement
December 2019	Schedule: Finish Slippage	Risk-adjusted RSD: 16 July 2020; accepted baseline: 31 March 2020	4.35%	Elevated (60), conditional	Package A baseline-duration proxy
December 2019	Schedule: Milestone Delay Rate	Thirteen selected milestones with later forecast dates	13/13 (100%)	Not formally scored	Selected-population diagnostic
December 2019	Change: Total Change Exposure	Approved: $292.762m; pending: $36.341m	11.85%	High (80)	Source-matched indicator
December 2019	Change: Pending Change Exposure	Pending changes: $36.341m	1.31%	Watch (30)	Source-matched indicator
December 2019	Change: Change Trend Growth	Change exposure compared with June 2019	0.17%	Not formally scored	Unequal reporting intervals; descriptive comparison only
December 2019	Contractor: Missed Milestones	Two missed milestones among two selected contractual milestones	2/2 (100%)	Not formally scored	Selected-population diagnostic; failures already realised

Panel B. East Side Access / Grand Central Madison

Reporting period	Domain and indicator	Historical inputs or	Recalculated value	IPCRF classification	Evidence status and interpretation

		calculation basis			
October 2018	Cost: EAC Variance	EAC: $11,133.3m; frozen 2014 budget: $10,177.8m	9.39%	Elevated (60), conditional	Frozen historical budget, not necessarily contemporaneous approved budget
October 2018	Schedule: Finish Slippage	Forecast: 21 February 2022; internal target: 12 February 2021; duration proxy: 7,104 days	5.26%	Elevated (60), conditional	Construction-start proxy paired with rebaselined finish
October 2018	Schedule: Programme Contingency Depletion	Original programme buffer: 994 days; remaining: 295 days	70.32%	Not formally scored	Descriptive programme-contingency measure; not certified CPM Float Erosion
October 2018	Constructability: Access/Outage Conflicts	Three narrative-derived constraints	3	Not formally scored	Complete open-issue register and deduplication unavailable
May 2019	Cost: EAC Variance	EAC: $11,133.3m; frozen budget: $10,177.8m	9.39%	Elevated (60), conditional	Carried-forward forecast; no demonstrated cost acceleration
May 2019	Schedule: Finish Slippage	Forecast: 14 February 2022; internal target: 12 February 2021	5.17%	Elevated (60), conditional	Construction-start proxy
May 2019	Schedule: Programme Contingency Depletion	Original programme buffer: 994 days;	69.62%	Not formally scored	Not equivalent to formal CPM Float Erosion

		remaining: 302 days			
May 2019	Constructability: Access/Outage Conflicts	Three narrative-derived constraints	3	Not formally scored	Descriptive count only
May 2019	Constructability: Interface Failure	One selected coordination-point comparison	1/1 (100%)	Not formally scored	Selected-population diagnostic; no project-wide handoff denominator
September 2019	Cost: EAC Variance	EAC: $11,133.3m; frozen budget: $10,177.8m	9.39%	Elevated (60), conditional	Carried-forward historical EAC
September 2019	Schedule: Finish Slippage	Forecast: 14 February 2022; internal target: 12 February 2021	5.17%	Elevated (60), conditional	Construction-start proxy
September 2019	Schedule: Programme Contingency Depletion	Original programme buffer: 994 days; remaining: 302 days	69.62%	Not formally scored	Descriptive programme-contingency measurement
September 2019	Constructability: Access/Outage Conflicts	Five narrative-derived constraints	5	Not formally scored	Complete issue population and open status not certified
September 2019	Constructability: Interface Failure	Fourteen unresolved later forecasts among 28 selected coordination points	14/28 (50.00%)	Not formally scored	Selected-population diagnostic; see Supplementary Table S3

Panel C. NYCT ADA station improvements

Reporting period	Domain and indicator	Historical inputs or calculation basis	Recalculated value	IPCRF classification	Evidence status and interpretation
July 2021	Cost: EAC Variance	Current budget: $170.4m; EAC: $169.0m	−0.82%	Low (10), conditional	Rounded contemporaneous values; cost-basis comparability requires qualification
July 2021	Schedule: Finish Slippage	Forecast and original completion target: July 2023	0.00%	Low (10)	Source-matched indicator
January 2022	Cost: EAC Variance	Reported budget: approximately $170m; EAC: $169m	−0.59%	Low (10), conditional	Rounded amounts; historical budget basis differs from later reports
January 2022	Schedule: Finish Slippage	Forecast and original completion target: July 2023	0.00%	Low (10)	Source-matched indicator
February 2023	Cost: EAC Variance	Current budget and EAC: $163m each	0.00%	Low (10)	Corrected same-report comparison; prior cross-basis 1.62% withdrawn
February 2023	Schedule: Finish Slippage	Forecast: 1 January 2024; original target: 1 July 2023; baseline duration: 1,064 days	17.29%	Critical (100)	Source-matched indicator; first day of forecast month assumed
October 2024	Cost: EAC Variance	Current budget: $163m; EAC: $168m	3.07%	Watch (30)	Source-matched indicator; distinct from $169m award budget
October 2024	Schedule: Finish Slippage	Forecast: 1 March 2025; original target: 1 July 2023; baseline duration: 1,064 days	57.24%	Critical (100)	Source-matched indicator; first day of forecast month assumed

Panel D. Canarsie Tube / L Project Tunnel Rehabilitation

Reporting period	Domain and indicator	Historical inputs or calculation basis	Recalculated value	IPCRF classification	Evidence status and interpretation
29 Sep 2019	Schedule: Finish Slippage	S2; April 2020 forecast preceded source-stated planned finish; negative slippage floored at zero	0.00%	Low (10)	**Scoreable.** Historical project-level snapshot; source described rehabilitation as approximately three months ahead of projected duration
25 Mar 2020	Schedule: Finish Slippage	S10; April 2020 forecast; no evidence of later-than-plan finish	0.00%	Low (10)	**Scoreable.** Historical project-level snapshot
26 Apr 2020	Schedule: Finish Slippage	S11; tunnel work completed ahead of schedule; negative outcome variance floored at zero	0.00%	Low (10)	**Outcome-state check only.** Later official material quantified completion as three months early
30 Jun 2020	Cost: Contingency Stress Ratio — ET040222 Shaft Excavation–1 Avenue	S3; MTA Contingency Index; 99% complete; EAC $17,335,502	0.34	Low (10)	**Component-level only.** No project-level aggregation rule applied
30 Jun 2020	Cost: MTA Cost Index — ET040222	S3; EAC/current project budget amount	0.97	—	**Context only; not substituted for IPCRF EAC Variance.** Current budget is not BAC
30 Jun 2020	Schedule: MTA Schedule Variance — ET040222	S3; change in schedule since prior Traffic Light Report	0 months	—	**Context only; not substituted for IPCRF Finish Slippage.** Not baseline-normalised

30 Jun 2020	Cost: Contingency Stress Ratio — ET050209 Mainline Track	S3; MTA Contingency Index; 100% complete; EAC $34,504,672	0.05	Low (10)	**Component-level only.** No project-level aggregation rule applied
30 Jun 2020	Cost: MTA Cost Index — ET050209	S3; EAC/current project budget amount	1.00	—	**Context only; not substituted for IPCRF EAC Variance**
30 Jun 2020	Schedule: MTA Schedule Variance — ET050209	S3; quarter-over-quarter measure	0 months	—	**Context only; not substituted for IPCRF Finish Slippage**
30 Jun 2020	Cost: Contingency Stress Ratio — ET060213 Tunnel Lighting	S3; MTA Contingency Index; 92% complete; EAC $50,423,338	0.00	Low (10)	**Component-level only.** No project-level aggregation rule applied
30 Jun 2020	Cost: MTA Cost Index — ET060213	S3; EAC/current project budget amount	1.00	—	**Context only; not substituted for IPCRF EAC Variance**
30 Jun 2020	Schedule: MTA Schedule Variance — ET060213	S3; quarter-over-quarter measure	0 months	—	**Context only; not substituted for IPCRF Finish Slippage**
30 Jun 2020	Cost: Contingency Stress Ratio — ET060219 Pump Room	S3; MTA Contingency Index; 98% complete; EAC $24,173,895	0.94	Watch (30)	**Component-level only.** No project-level aggregation rule applied
30 Jun 2020	Cost: MTA Cost Index — ET060219	S3; EAC/current project budget amount	1.00	—	**Context only; not substituted for IPCRF EAC Variance**
30 Jun 2020	Schedule: MTA Schedule Variance — ET060219	S3; quarter-over-quarter measure	0 months	—	**Context only; not substituted for IPCRF Finish Slippage**
30 Jun 2020	Cost: Contingency Stress Ratio — ET080211 Signals	S3; MTA Contingency Index; 96% complete;	0.21	Low (10)	**Component-level only.** No project-level aggregation rule applied

		EAC $26,109,726			
30 Jun 2020	Cost: MTA Cost Index — ET080211	S3; EAC/current project budget amount	0.71	—	**Context only; not substituted for IPCRF EAC Variance**
30 Jun 2020	Schedule: MTA Schedule Variance — ET080211	S3; quarter-over-quarter measure	0 months	—	**Context only; not substituted for IPCRF Finish Slippage**
30 Jun 2020	Cost: Contingency Stress Ratio — ET090211 Two Circuit Breaker Houses	S3; MTA Contingency Index; 97% complete; EAC $36,178,677	0.54	Low (10)	**Component-level only.** No project-level aggregation rule applied
30 Jun 2020	Cost: MTA Cost Index — ET090211	S3; EAC/current project budget amount	0.99	—	**Context only; not substituted for IPCRF EAC Variance**
30 Jun 2020	Schedule: MTA Schedule Variance — ET090211	S3; quarter-over-quarter measure	0 months	—	**Context only; not substituted for IPCRF Finish Slippage**
30 Jun 2020	Cost: Contingency Stress Ratio — ET090212 Power/Communication Cable and Ducts	S3; MTA Contingency Index; 96% complete; EAC $283,790,379	0.16	Low (10)	**Component-level only.** No project-level aggregation rule applied
30 Jun 2020	Cost: MTA Cost Index — ET090212	S3; EAC/current project budget amount	0.81	—	**Context only; not substituted for IPCRF EAC Variance**
30 Jun 2020	Schedule: MTA Schedule Variance — ET090212	S3; quarter-over-quarter measure	0 months	—	**Context only; not substituted for IPCRF Finish Slippage**
30 Jun 2020	Cost: Contingency Stress Ratio — ET090309 Mitigation: Power/Communication Cable and Ducts	S3; MTA Contingency Index; 100% complete; EAC $49,704,193	0.00	Low (10)	**Component-level only.** No project-level aggregation rule applied

30 Jun 2020	Cost: MTA Cost Index — ET090309	S3; EAC/current project budget amount	0.49	—	**Context only; not substituted for IPCRF EAC Variance**
30 Jun 2020	Schedule: MTA Schedule Variance — ET090309	S3; quarter-over-quarter measure	0 months	—	**Context only; not substituted for IPCRF Finish Slippage**
1 Jan 2025*	Cost: Later official cost outcome	S6; later MTA capital-plan statement	$100 million under budget	—	**Outcome-state check only.** Not converted to EAC Variance because public cost figures use different scope bases

Note. IPCRF = Integrated Project Controls and Risk Forecasting Framework; BAC = budget at completion; EAC = estimate at completion; CPI = cost performance index; EV = earned value; RSD = revenue-service date; CPM = critical path method; NYCT = New York City Transit. Values are rounded for presentation; full-precision calculations are retained in the analytical registers. Source-matched indicators use compatible reported inputs and measurement definitions. Conditional indicators depend on stated assumptions regarding baseline, scope, normalisation or reporting interval. Selected-population diagnostics do not represent complete project-wide indicator populations. Descriptive and contextual measurements are not assigned formal IPCRF risk scores. The withdrawn Dulles Cost Trend Acceleration calculation is retained solely to document its exclusion. Across 12 primary project-level observations and 23 IPCRF indicators, 276 potential project-level indicator-period cells were considered; 227 remained unavailable after reconciliation. Thirty-three project-level classifications were retained. Case 4 component-level and outcome-state records are reported separately and do not enter the 276-cell denominator. No complete six-domain PDRS or FCS was calculated for any case.

Supplementary Table S2. Source verification, temporal provenance, baseline and outcome register

Panel A. Historical reporting sources

ID	Project	Historical source / period	Main information used	Analytical role	Key qualification
N1	NYCT	MTA Capital Program Committee, July 2021	Budget, EAC, July 2023 completion forecast	Primary observation	Original public-posting date not independently established
N2	NYCT	MTA Capital Program Committee, January 2022	Budget, EAC, completion forecast and IEC schedule concerns	Primary observation	Narrative concerns were not converted into quantitative indicators
N3	NYCT	MTA Capital Program Committee, February 2023	Current budget, EAC and revised completion forecast	Primary observation	Retrospectively reported station information not treated as separate earlier source

N4	NYCT	MTA Capital Program Committee, October 2024	Current budget, EAC and revised completion forecast	Primary observation	Contemporary budget basis retained separately from earlier cost references
D1	Dulles	FTA PMOC, March 2019	Budget/EAC, EV, actual cost, contingency, changes and contractual milestones	Primary observation	Reporting period not equivalent to verified first public availability
D2	Dulles	FTA PMOC, June 2019	Cost, change exposure, milestones and risk-adjusted RSD	Primary observation	Variables have differing underlying data cut-offs
D3	Dulles	FTA PMOC, December 2019-labelled report	Cost, schedule, changes and contractor milestones	Primary observation	Finalised February 2020; principal data current through November 2019
D-S1	Dulles	FTA PMOC, February–May 2019	Change exposure and compatible schedule information	Sensitivity analysis	Used to test whether qualifying classifications preceded selected primary observations
D-S2	Dulles	FTA PMOC, August–November 2019	Change exposure and compatible project-control information	Sensitivity analysis	Used to assess persistence, peak severity and recovery
E1	East Side Access	FTA PMOC, October 2018	EAC, internal RSD, programme contingency and construction constraints	Primary observation	Schedule information reflects earlier IPS data cut-off
E2	East Side Access	FTA PMOC, May 2019	Cost, schedule contingency, access and coordination conditions	Primary observation	Schedule information reflects April 2019 IPS data
E3	East Side Access	FTA PMOC, September 2019	Cost, schedule, access conditions and coordination-point data	Primary observation	Relevant schedule/interface information reflects August 2019 IPS data
E-S1	East Side Access	FTA PMOC, May–June 2018	Earlier target RSD and April 2018 cost reassessment	Sensitivity analysis	Established earlier compatible schedule and cost classifications
E-S2	East Side Access	Additional compatible	Target RSD and programme-	Sensitivity analysis	Used to assess persistence and severity

		PMOC reports, 2018–2019	contingency information		

Panel B. Baseline and outcome verification

Project	Reference variable	Date/value	Status	Methodological treatment	Source
NYCT	Historical open-data budget	$160.409 million	Historical cost reference	Retained separately from contemporaneous budget/EAC comparisons	MTA project register
NYCT	Original programme completion	1 July 2023	Original schedule baseline	Principal Finish Slippage comparator	MTA (2023b)
NYCT	February 2023 programme forecast	January 2024	Historical forecast	Used in Finish Slippage reconstruction	MTA (2023b)
NYCT	Later estimated completion	30 June 2026	Later forecast	Not treated as verified actual project-wide completion	MTA Capital Dashboard
Dulles	Original revenue-service date	31 January 2019	Earlier baseline	Preserved separately from accepted revised baseline	FTA PMOC records
Dulles	Accepted revenue-service baseline	31 March 2020	Rebaselined target	Principal historical schedule comparator	FTA PMOC records
Dulles	Risk-adjusted revenue-service forecast	16 July 2020	Historical forecast	Used conditionally; not treated as accepted reconciled master schedule	FTA PMOC records
Dulles	Actual revenue-service commencement	15 November 2022	Observed outcome	Retrospective outcome comparison only	WMATA (2022)
East Side Access	June 2014 re-plan budget	$10,177.8 million	Frozen historical cost reference	Retained separately from later approved-budget definitions	FTA (2014)

East Side Access	Internal revenue-service target	12 February 2021	Internal target	Principal comparator for internal schedule deterioration	FTA (2014)
East Side Access	Late/replanned RSD benchmark	13 December 2022	Later schedule benchmark	Retained separately from internal target	FTA (2014)
East Side Access	Limited shuttle-service commencement	25 January 2023	Partial-service outcome	Not treated as full-service commencement	MTA (2023a)
East Side Access	Full-service commencement	27 February 2023	Observed outcome	Principal operating outcome	MTA (2023c)
Canarsie	2019 rehabilitation phase	April 2019–Summer 2020	Published programme window	Scope/timing context only	S1
Canarsie	29 September 2019 forecast	April 2020; approximately three months ahead of projected duration	Historical forecast	Project-level Finish Slippage snapshot; negative difference floored at zero	S2
Canarsie	25 March 2020 forecast	April 2020	Historical forecast	Second project-level Finish Slippage snapshot	S10
Canarsie	Tunnel completion	26 April 2020; later described as three months early	Observed outcome	Outcome check only	S6, S11
Canarsie	Original repair-contract narrative	Approximately $492 million	Contract-scope cost reference	Retained separately; not treated as interchangeable with broader programme values	S8
Canarsie	P36437 current amount in 2023 report	$512,566,795	Current contract amount	Not treated as original award/BAC	S5
Canarsie	P36437 life-to-date expenditures	$499,629,415.56	Contract expenditure reference	Scope-specific contextual evidence	S5
Canarsie	Broader Canarsie	$781.0 million	Broader programme amount	Not used as tunnel BAC	S4

	restoration/core-capacity amount				
Canarsie	Later cost outcome	$100 million under budget	Observed favorable outcome	Outcome check; exact EAC Variance not calculated because a matching same-scope denominator was unavailable	S6

Panel C. Case Study 4 source register

ID	Public source	Back-test use	Official source
S1	MTA — *L Project brochure* (July 2019)	Phase 2 timing and project scope	https://files.mta.info/s3fs-public/2019-07/19_112_L_Brochure_%20July%201%2C%202019.pdf
S2	FTA — *TAMNews* (December 2019)	September 2019 progress; first tube ahead/on budget; overall rehabilitation expected approximately three months early	https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/regulations-and-programs/asset-management/146541/tamnewsdecember-2019_0.pdf
S3	MTA — Capital Program Oversight Committee, 2Q2020 Traffic Light Report	Component-level EAC, completion, Contingency Index, Cost Index, Schedule Variance and metric definitions	https://www.mta.info/document/20211
S4	MTA — *2020 Annual*	Completion scope and	https://www.mta.info/document/78921

	Report to the Governor	broader combined Canarsie restoration/core-capacity amount	
S5	MTA — PARIS 2023 All Agency Consolidation Report	P36437 contract metadata, amount, expenditures and end date	https://www.mta.info/document/135346
S6	MTA — *2025–2029 Capital Plan*	Later official outcome statement: three months ahead and $100 million under budget	https://www.mta.info/document/151266
S7	MTA — capital-savings press release	Later confirmation that rehabilitation was completed ahead of schedule and under budget	https://www.mta.info/press-release/mta-reports-details-of-nearly-300-million-capital-savings
S8	New York City Council — transportation oversight record	Original repair-contract narrative, completion incentive and liquidated-damages structure	https://legistar.council.nyc.gov/View.ashx?GUID=BA26975E-C43F-4CEE-B621-96B987D6F87A&ID=9346250&M=F
S9	FTA — Canarsie Reevaluation Checklist	Revised rehabilitation approach and project/contract context	https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/regulations-and-guidance/environmental-programs/133456/reevalution-canarsie-checklist.pdf
S10	MTA — Committee/Board record,	Near-completion forecast; tunnel	https://www.mta.info/document/16626

	25 March 2020	completion still anticipated in April 2020	
S11	Official New York State/Governor completion announcement	Contemporaneous completion statement: project completed ahead of schedule and under budget	https://www.youtube.com/watch?v=ZTOfixtc21A

Note. FTA = Federal Transit Administration; MTA = Metropolitan Transportation Authority; NYCT = New York City Transit; PMOC = Project Management Oversight Contractor; IEC = Independent Engineering Consultant; IPS = Integrated Project Schedule; RSD = revenue-service date; EAC = estimate at completion. The primary harmonised project-level dataset contains 12 observations: four NYCT, three Dulles, three East Side Access and two Canarsie historical snapshots. Additional Dulles and East Side Access records were examined separately in the reporting-period sensitivity analysis. Case 4 component-level 2Q2020 observations and outcome-state checks are not counted as additional project-level reporting periods. Reporting period, underlying data cut-off, report finalisation and first public availability are treated as distinct temporal variables. Original public-release dates could not be independently verified for all records; calculated intervals are therefore not interpreted as validated public-warning lead times.

Supplementary Table S3: Reconstructed coordination-point activities from the September 2019 East Side Access PMOC report

No.	Activity ID	Activity description	May IPS date	August IPS date	Change (days)	Event classification
1	CP-CM007-250	Install GCT 4 Switches	1 May 2019	1 Aug 2019	+92	Boundary
2	CP-VM014-420	Escalator and Elevator Deliveries for CM007	1 May 2019	4 Apr 2019 A	−27	Other
3	CP-CH061A-20	Install Equipment in Tunnel A	1 May 2019	1 Mar 2019 A	−61	Other
4	CP-CM014B-360	MS13 Turnover to CM007 EL-2	2 May 2019	2 Aug 2019	+92	Unresolved later forecast
5	CP-CH057D-10	Turnover Track B/C Area	6 May 2019	10 Mar 2019 A	−57	Other

6	CP-VS086-20	Procure Signals GCT-6	6 May 2019	21 Dec 2018 A	−136	Other
7	CP-VQ033-10-CQ033	MID1 CIL Delivery	6 May 2019	10 Apr 2019 A	−26	Other
8	CP-CM007-110	CIA 19C CS179 EC Platform Conduits	9 May 2019	7 Aug 2019	+90	Unresolved later forecast
9	CP-CM007-060	CIA 17B CS179 Civil Electrical	9 May 2019	5 Jul 2019 A	+57	Completed late
10	CP-CM007-140	CIA 20C CS179 WC Upper Conduits	10 May 2019	2 Aug 2019	+84	Unresolved later forecast
11	CP-VS086-30	Procure Signals GCT-5	14 May 2019	6 Feb 2019 A	−97	Other
12	CP-VQ033-20-CQ033	MID2 CIL Delivery	15 May 2019	15 May 2019 A	0	Other
13	CP-VM014-440	Escalator/Elevator Deliveries Complete	16 May 2019	16 Apr 2019 A	−30	Other
14	CP-VS086-50	Procure Signals GCT4	21 May 2019	30 May 2019	+9	Unresolved later forecast
15	CP-CM007-100	CIA 19B CS179 EC Mezzanine Conduits	22 May 2019	12 Aug 2019	+82	Unresolved later forecast
16	CP-CM007-370	Energise/Commission US1/US2	22 May 2019	22 Aug 2019	+92	Unresolved later forecast
17	CP-CM007-180	Install GCT 5 Switches	28 May 2019	16 Aug 2019	+80	Unresolved later forecast
18	CP-CM007-050	CIA 17A CS179 Civil/Electrical	29 May 2019	5 Jul 2019 A	+37	Completed late
19	CP-CM014B-110	MS18A Conduits Zone 1	30 May 2019	22 Aug 2019	+84	Unresolved later forecast
20	CP-CM014B-190	MS18A Conduits Zone 2	30 May 2019	22 Aug 2019	+84	Unresolved later forecast

21	CP-CM007-130	CIA 20B CS179 WC Conduits and Cable	31 May 2019	5 Sep 2019	+97	Unresolved later forecast
22	CP-CH058A-20	Demolition GO2 Substation	14 Jun 2019	7 Dec 2018 A	−189	Other
23	CP-CM007-070	CIA 18A CS179 Civil/Electrical	19 Jun 2019	21 Jun 2019 A	+2	Completed late
24	CP-CM014B-350	MS18B Conduits Zones 4–5	25 Jun 2019	26 Sep 2019	+93	Unresolved later forecast
25	CP-CM014B-270	MS18B Conduits Zone 3	25 Jun 2019	26 Sep 2019	+93	Unresolved later forecast
26	CP-CM007-20	Complete Track CO2	28 Jun 2018	13 Sep 2019	+442	Unresolved later forecast; baseline-year anomaly
27	CP-VS086-40	Procure Signals GCT3	10 Jul 2019	25 Mar 2019 A	−107	Other
28	CP-CM014B-30	MS20B Zone 2 Architectural Finishes	30 Jul 2019	7 Aug 2019	+8	Unresolved later forecast

Supporting calculation for Supplementary Table S3

Analytical definition	Numerator	Denominator	Descriptive proportion
Unresolved later forecasts only	14	28	50.00%
Unresolved later forecasts and completed-late activities	17	28	60.71%
Unresolved later forecasts, completed-late activities and boundary item	18	28	64.29%

Note. IPS = Integrated Project Schedule. The May and August headings identify the respective schedule versions. The suffix A identifies an actual completion date in the source register. Positive changes indicate a later date in the August schedule relative to the May schedule. The narrow definition identifies 14 unresolved later forecasts among 28 selected coordination points. Three activities were completed late, and one unactualised activity was forecast exactly on the August schedule-data date. Activity CP-CM007-20 is retained as transcribed despite the apparent 2018 baseline-year anomaly. These selected activities do not establish the complete population of project-wide interface handoffs and must not be presented as a formally certified IPCRF Interface Failure Rate. No equivalent project-wide interface-event register was available for Case 4; therefore, Canarsie is not represented in Supplementary Table S3. Only the 14/28 definition is used in the Results; the broader 17/28 and 18/28 definitions are sensitivity descriptions rather than validated project-wide rates.

Supplementary Table S4: Formula, provisional threshold and weight register for IPCRF indicators classified in the primary analysis

Domain	Indicator	Formula	Low (10)	Watch (30)	Elevated (60)	High (80)	Critical (100)	Within-domain weight
Cost	EAC Variance (%)	[(EAC − BAC) / BAC] × 100	≤2	>2–5	>5–10	>10–15	>15	40%
Cost	Cost Performance Index (CPI)	EV / Actual Cost	≥1.00	0.95–0.999	0.90–0.949	0.85–0.899	<0.85	25%
Cost	Contingency Stress Ratio	% contingency consumed / % physical progress	≤0.80	>0.80–1.00	>1.00–1.25	>1.25–1.50	>1.50	20%
Schedule	Finish Slippage (%)	[max(0, Current Finish − Baseline Finish) / Baseline Duration] × 100	≤1	>1–3	>3–7	>7–12	>12	45%
Change	Total Change Exposure (%)	[(Approved Change + Pending Change) / BAC] × 100	≤2	>2–5	>5–10	>10–15	>15	35%
Change	Pending Change Exposure (%)	(Pending Change / BAC) × 100	≤1	>1–3	>3–5	>5–8	>8	25%

Note. BAC = budget at completion; EAC = estimate at completion; EV = earned value. Thresholds and within-domain weights are those specified in IPCRF Version 1.2 and were applied unchanged in the retrospective historical back-test. Version 1.2 identifies the thresholds as provisional starter values rather than empirically calibrated universal cut-points. The table covers indicator types contributing to the 33 retained project-level indicator classifications. The eight Case 4 component-level Contingency Stress mappings use the same frozen thresholds but are reported separately from the project-level classification count. Cost Trend Acceleration was excluded because its consecutive-period requirement was not satisfied. Milestone-delay proportions, contractor milestone failures, programme-contingency depletion, access/outage observations, coordination-point measures and non-equivalent MTA contextual metrics were retained as diagnostics or descriptive/contextual evidence and are not included as formal project-level classifications. No formal domain, PDRS or FCS score was inferred merely from the presence of individual indicator classifications.

Supplementary Table S5. Reporting-period sensitivity analysis for Dulles Phase 2 and East Side Access

Project	Additional report/underlying period	Indicator assessed	Reconstructed result	Classification/status	Effect on primary finding
Dulles	January 2019	EAC Variance	0.00%	Low	Confirms cost stability before March primary observation
Dulles	February 2019	Total Change Exposure	11.57%	High (80)	Moves earliest observed High change exposure earlier than March
Dulles	February 2019	Pending Change Exposure	2.95%	Watch (30)	Shows Pending Change had not yet crossed Elevated
Dulles	5 March 2019 schedule status	Finish Slippage	6.02%	Elevated (60), conditional	Moves earliest exact-date conditional schedule classification earlier than June
Dulles	April 2019	Total Change Exposure	11.81%	High (80)	Confirms persistence
Dulles	April 2019	Pending Change Exposure	3.04%	Elevated (60)	Earliest compatible Elevated Pending Change observation
Dulles	May 2019	Total Change Exposure	11.82%	High (80)	Confirms persistence
Dulles	May 2019	Pending Change Exposure	3.03%	Elevated (60)	Confirms persistence
Dulles	July data reported August 2019	Total Change Exposure	12.38%	High (80)	Greater magnitude but same band
Dulles	September 2019	Total Change Exposure	12.42%	High (80)	Same severity band
Dulles	September 2019	Pending Change Exposure	2.63%	Watch (30)	Demonstrates return from Elevated to Watch
Dulles	October 2019	Total Change Exposure	12.47%	High (80)	Highest reviewed magnitude; still High
Dulles	October 2019	Pending Change Exposure	2.61%	Watch (30)	Confirms later reduction

Dulles	November 2019	Total Change Exposure	11.84%	High (80)	High condition persisted
Dulles	November 2019	Pending Change Exposure	1.98%	Watch (30)	Confirms lower unresolved exposure
East Side Access	May 2018; 1 April IPS data	Finish Slippage	5.26%	Elevated (60), conditional	Moves earliest observed schedule classification earlier than October 2018
East Side Access	June 2018	EAC Variance	9.39%	Elevated (60), conditional	Moves earliest cost evidence earlier than October 2018
East Side Access	August/September 2018	Programme Contingency Depletion	≈70%	Descriptive only	Shows schedule-buffer deterioration preceded October primary record
East Side Access	November 2018	Finish Slippage	5.26%	Elevated (60), conditional	Confirms persistence
East Side Access	January–March 2019	Finish Slippage	≈5.17%	Elevated (60), conditional	Confirms persistence after target moved to 14 February 2022
East Side Access	June–August 2019	Finish Slippage	≈5.17%	Elevated (60), conditional	Confirms persistence
East Side Access	October 2019	Finish Slippage	5.56%	Elevated (60), conditional	Numerical worsening but no band change
East Side Access	October 2019	Programme Contingency Depletion	≈72.4%	Descriptive only	Greater depletion; still not formal CPM Float Erosion
East Side Access	November 2019	Finish Slippage	≈5.26%	Elevated (60), conditional	No recovery to Low/Watch

Note. The sensitivity analysis evaluates whether compatible monthly records omitted from the selected primary observations materially changed the earliest observed qualifying classification, persistence, recovery or maximum severity. These observations are not added to the 12 primary project-level observations. The additional Dulles and East Side Access records identify earlier qualifying observations in several instances but do not increase the maximum classification bands underpinning the main cross-case interpretation. "Earliest observed" refers only to the earliest compatible evidence located within the reviewed records and must not be interpreted as the true onset of deterioration or first public disclosure. No separate reporting-period sensitivity series was constructed for Case 4. Its analysis is restricted to the selected historical project-level snapshots, component-level 2Q2020 evidence and separately identified outcome checks.