Journal of
Cardiovascular Development and Disease

MDPI

Review

Defining the Ideal Operation: Controversies and Evolving Paradigms in Textbook Outcome for Adult Cardiac Surgery

Dimitrios E. Magouliotis [1,2,*], Vasiliki Androutsopoulou [3], Prokopis-Andreas Zotos [3], Serge Sicouri [1], Basel Ramlawi [1,4], Dimitrios Zacharoulis [5] and Andrew Xanthopoulos [2]

[1] Department of Cardiac Surgery Research, Lankenau Institute for Medical Research, Wynnewood, PA 19096, USA; sicouris@mlhs.org (S.S.); ramlawib@mlhs.org (B.R.)
[2] Department of Cardiology, University of Thessaly, Biopolis, 41110 Larissa, Greece; andrewvxanth@gmail.com
[3] Department of Cardiothoracic Surgery, University of Thessaly, Biopolis, 41110 Larissa, Greece; androutsopoulouvasiliki@uth.gr (V.A.); zotospro@hotmail.com (P.-A.Z.)
[4] Department of Cardiac Surgery, Lankenau Medical Center, Wynnewood, PA 19096, USA
[5] Department of Surgery, University of Thessaly, Biopolis, 41110 Larissa, Greece; zacharoulis@uth.gr
* Correspondence: magouliotisd@mlhs.org

Abstract

Textbook outcome (TO), an all-or-none composite describing the ideal perioperative course, has entered adult cardiac surgery. This narrative review, based on a structured PubMed search (9 November 2026; 43 records), compares adult cardiac TO definitions, uses transplant, congenital and aortic definitions as methodological comparators, and examines four controversies: definition heterogeneity, high achievement rates and statistical reliability, positioning against existing quality metrics, and patient-centeredness and equity. Two registry analyses report TO in 86.0% of 963,775 elective coronary and valve admissions and in 87.1% of 3026 mitral operations, with component sets that differ in length of stay, readmission and time horizon. In the national cohort, prolonged hospitalization was present in 58.6% of failures, and hospital TO rates correlated weakly with in-hospital mortality and failure to rescue (FTR). Two coronary series with a stricter definition report 59.1% to 81.9%. No adult cardiac study has reported a validated TO risk model, hospital-level reliability, or a comparison with the Society of Thoracic Surgeons composite score. We propose, as a preliminary framework requiring consensus and external validation, a registry-based core TO set, supplementary patient-centered and 1-year durability indicators reported separately, paired reporting of TO and FTR, and a research agenda to establish incremental value.

Keywords: textbook outcome; failure to rescue; composite quality metric; cardiac surgery; quality improvement; benchmarking; risk adjustment; patient-centered outcomes

Academic Editor: Jacek Kubica

Received: 20 October 2026
Revised: 13 November 2026
Accepted: 20 November 2026
Published: 22 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Operative mortality after adult cardiac surgery is now low. The Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database reports an overall mortality of 2.6% across isolated coronary artery bypass grafting (CABG), aortic valve replacement with or without CABG and mitral valve repair or replacement with or without CABG, and 0.9% among patients who have none of the four complications used in the STS failure to rescue (FTR) metric [1]. EuroSCORE II, built on 22,381 patients from 43 countries, recorded an

observed mortality of 3.9% against a predicted 5.8% from the original additive model, which documents how far outcomes had moved in 15 years [2]. An event that occurs in fewer than 3 of every 100 patients gives little statistical power to separate centers at ordinary volumes, and it says nothing about the 97 patients who survive [3]. Quality measurement in cardiac surgery has therefore expanded over two decades from mortality alone to risk-adjusted morbidity, process measures and composite star ratings [4–7]. The STS 2018 risk models include separate endpoints for stroke, renal failure, prolonged ventilation, reoperation, deep sternal wound infection and prolonged or short length of stay, and a composite of major morbidity or mortality [8,9]. Each of these endpoints is a count of something that went wrong. None of them describes the patient who had the operation that every surgeon would want for a relative: discharged alive, on time, without a complication, without a return to the operating room and without readmission.

Textbook outcome (TO) is the metric that describes that patient. It was introduced in Dutch colorectal surgery in 2013 as an all-or-none composite of desirable events and was applied to esophagogastric cancer surgery and aneurysm repair in 2017, pancreatic surgery in 2020 and lung resection in 2021 [10–14]. In 2025, our group published a systematic review of TO across cardiothoracic surgery, which concluded that benchmarking of TO in adult cardiac procedures remained limited and non-standardized and called for harmonized, procedure-specific core TO sets aligned with registry fields [15]. The call for core sets remains unmet, but the evidence base has grown. Since 2024, a national analysis of 963,775 elective CABG and valve admissions, a Netherlands Heart Registration (NHR) analysis of 3026 mitral operations, two single-center coronary series, two claims-based disparity analyses, a Fontan series and several aortic registry studies have each applied a TO definition [16–25]. The question for cardiac surgery is no longer whether TO can be measured but whether it should be measured as currently defined, how it should be adjusted, and what it adds to the metrics that are already in place. This review takes those questions in turn. Its primary scope is adult cardiac surgery. Heart transplant, congenital and aortic TO studies are included as methodological comparators only: they show how choices about length-of-stay thresholds, time horizons and readmission behave in registries, and no rate or association from those populations is generalized to adult cardiac surgery. The subject is treated as a set of controversies rather than as a catalog of results. The review methodology is reported in Section 2, and Section 10 separates what the published studies establish from what we infer and from what remains to be tested.

2. Review Methodology

This is a narrative review supported by a structured and reproducible search, and its reporting was guided by the Scale for the Assessment of Narrative Review Articles [26]. PubMed (MEDLINE) was the only database searched. A preliminary search on 4 October 2026 was restricted to the Title field. It is superseded by the search reported here, which was run on 9 November 2026 in the Title/Abstract field with the string (“textbook outcome”[tiab] OR “textbook outcomes”[tiab] OR “text-book outcome”[tiab]) AND (cardiac[tiab] OR cardiothoracic[tiab] OR heart[tiab] OR coronary[tiab] OR CABG[tiab] OR valve[tiab] OR mitral[tiab] OR aortic[tiab] OR aneurysm[tiab] OR congenital[tiab] OR Norwood[tiab] OR Fontan[tiab] OR “Cardiac Surgical Procedures”[MeSH] OR “Heart Transplantation”[MeSH] OR “Aortic Aneurysm”[MeSH]), with no date or language limit. The TO terms alone returned 676 records and the combined string returned 43, all in English. The hyphenated variant retrieved no additional records. Titles and abstracts of the 43 records were screened and assigned to the categories that follow, and full texts were consulted where available. Original studies that applied a TO definition to adult CABG, valve or combined cardiac operations formed the primary evidence (6 records). Original TO studies in heart transplantation (2 records), congenital heart surgery (3 records) and

open or endovascular aortic repair (10 records) were retained as methodological comparators. Claims-based cohorts that pooled CABG with non-cardiac operations (11 records) were used only as context for the equity discussion; 2 records were reviews, and 9 records concerned neither cardiac nor aortic surgery and were excluded. This search identified studies that the preliminary search had missed, including four adult cardiac studies in which TO was not named in the title. Screening and classification were performed by the authors responsible for the literature search, without independent duplicate screening or agreement statistics. Table S1 lists all 43 records with their classification, so that the selection can be reproduced.

For each included study, we extracted the data source, cohort, TO components, length-of-stay criterion, ascertainment window, TO rate, component-level failure data and the outcomes tested for association with TO, and we kept components and associated outcomes in separate fields. During revision, every numerical estimate in the text, tables and figures was rechecked against its primary source, using the full text where it was available and the published abstract otherwise. Where a study reported a numerator and denominator without a confidence interval, we calculated a 95% interval by the Wilson method and labeled it as calculated. The methodological literature on composite measures, hierarchical modeling and reliability, prediction model development, FTR, days at home, patient-reported outcome measures and social risk adjustment was identified by targeted, non-systematic searching and from the reference lists of key papers, and was selected for relevance to the four controversies and not exhaustively. Registry definitions were taken from the STS Adult Cardiac Surgery Database training manual for data version 4.20.2 and from the European Association for Cardio-Thoracic Surgery (EACTS) Adult Cardiac Database data specification, version 3.3 [27,28]. We did not register a protocol, search other databases or the gray literature for TO studies, assess risk of bias, pool estimates or test for publication bias, and no patients or members of the public were involved in the design of the review or of the proposed framework. These choices limit the conclusions and are discussed in Section 11. All references were verified against their PubMed records, except the two registry documents, which were verified at source.

3. Conceptual Foundations

3.1. From Structure, Process and Outcome to the All-or-None Composite

Donabedian's triad of structure, process and outcome remains the organizing framework for quality measurement, and TO is an outcome measure in the strict sense: it records what happened to the patient, not what the hospital has or does [29]. What distinguishes TO from the outcome measures that preceded it is the scoring rule. Nolan and Berwick argued in 2006 that all-or-none measurement raises the bar on performance for three reasons: it reflects the patient's interest in receiving every element of good care rather than most of them, it encourages a systems view of care because a failure anywhere in the pathway is counted, and it is more sensitive to improvement than a mean of component rates [30]. TO applies this logic to outcomes. A patient either has the entire ideal course or does not. A hospital's TO rate is the proportion of its patients who had it. Because the rule is conjunctive, the TO rate can never exceed the success rate of any single component and cannot rise as components are added, which is the property that gives TO a wide range in high-morbidity procedures and a high, compressed range in low-morbidity ones.

The method also carries an interpretive advantage that the original Dutch authors emphasized. A single-indicator approach forces a reader to combine mortality, complications, reinterventions, length of stay and readmission mentally, and the combination differs from reader to reader. TO performs the combination explicitly and reports one number, with the decomposition available underneath it [10,11]. The empirical composite

measures developed by Dimick and Staiger reach a related conclusion from a different direction: combining several outcome domains and adjusting for reliability explains more hospital-level variation and predicts future performance better than any single risk-adjusted rate [31,32]. The difference is that empirical composites weight components statistically, whereas TO weights them equally and conjunctively, which is simpler to communicate and easier to game by dropping components that a center performs poorly on.

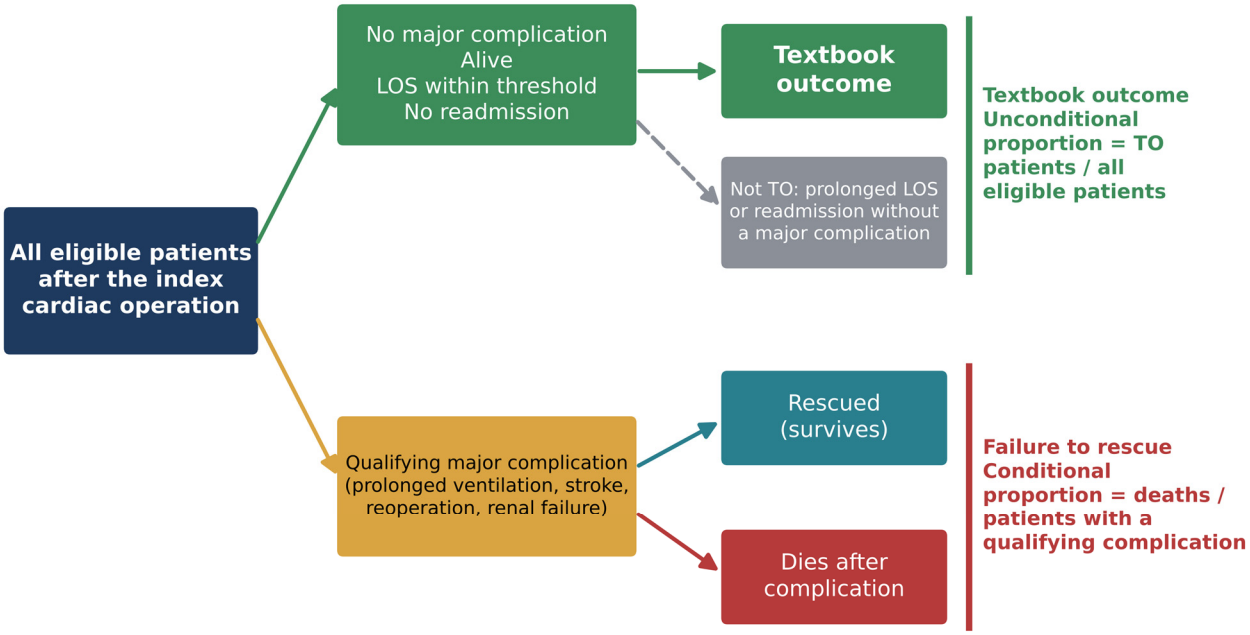
The all-or-none rule also has costs that the composite-measure literature has documented. First, the choice of aggregation method changes the answer. When five scoring methods, all-or-none among them, were applied to the same primary care data, 6 of 16 practices moved between the top and bottom quartiles depending on the method, and a review of composite construction found that provider rankings may be sensitive to such choices [33,34]. Second, a conjunctive composite is driven by its most frequent component, which is usually not its most severe. In 114 cardiovascular trials with composite endpoints, the components that mattered most to patients had lower event rates and smaller treatment effects than the minor components [35]. Applied to TO, a 15-day stay and a disabling stroke count identically, and a patient who fails one component is indistinguishable from a patient who fails five. Third, components are correlated, because a stroke or renal failure usually prolongs the stay, so the number of failed components carries information that the binary result discards. These properties do not invalidate TO. They mean that a TO rate is interpretable only with its component profile and with sensitivity analyses that test whether conclusions survive a different scoring rule, which we specify in Sections 6 and 10.

3.2. Textbook Outcome and Failure to Rescue: Complementary Measures with Different Denominators

The second conceptual anchor comes from the FTR literature. Silber showed in 1992 that hospital characteristics explained death after a complication better than they explained the complication itself [36]. Ghaferi, Birkmeyer and Dimick showed in 2009 that hospitals with high and low mortality after inpatient surgery had similar complication rates and differed mainly in how often a complicated patient died [37]. The cardiac surgical data are consistent. In the Michigan collaborative, complication rates ranged from 19.1% to 22.9% across low-, medium- and high-mortality centers, while FTR ranged from 6.6% to 13.5%, and a later analysis of 62,450 operations in the same 33 hospitals found FTR of 8.3%, 10.0% and 12.7% across mortality terciles [38,39]. In the STS database, complication rates after CABG rose modestly from 11.4% to 15.7% across mortality terciles while FTR more than doubled from 6.8% to 13.9%, and the STS Quality Measurement Task Force has since developed a risk-adjusted FTR metric with a c-statistic of 0.806 [1,40]. We have reviewed the FTR concept for cardiovascular surgery and for lung resection elsewhere [41,42].

TO and FTR are computed on different denominators and respond to different mechanisms, and we use that difference as a working framework for paired reporting, not as an established single construct (Figure 1). TO is an unconditional proportion: every eligible patient is in the denominator, and the rate is driven by complication avoidance, discharge practice and readmission. FTR is a conditional proportion: only patients who develop a qualifying complication are in the denominator, and the rate reflects recognition and response once a complication has occurred. Three consequences follow. First, the two measures are arithmetically linked but not redundant. With an identical complication list, every patient in the FTR denominator has already failed TO, so FTR describes the severity of a subset of TO failures and carries no information about patients who fail TO on length of stay or readmission alone. Second, the FTR denominator is selected by the hospital's own performance and coding. A hospital that prevents minor complications is left with a

smaller and sicker complicated group and can show a higher FTR with fewer deaths, and a hospital that records complications more completely enlarges its denominator and lowers its FTR without any change in care. In a case-note validation of 661 aneurysm repairs, complications were coded in 19.2% of administrative records and documented in 41.5% of case notes, and the reliability of FTR as a hospital measure was 0.32 in a Medicare analysis of 1567 hospitals [43,44]. Third, the two measures can diverge. A center with early discharge and weak intensive care can post a high TO rate and a high FTR rate; a center with cautious discharge practice can lose TO on length of stay while its complication and rescue performance are excellent; a low-risk center can post a high TO rate while failing to rescue the few patients who deteriorate. In the one adult cardiac cohort that examined both measures, hospital TO rates had minimal correlation with hospital FTR rates ($r = -0.09$), which suggests that divergence is common [16]. Whether the discordance of an individual hospital is stable from year to year and related to modifiable structures has not been studied.



The two measures have different denominators. Patients who fail TO on length of stay or readmission alone are not in the FTR denominator. A center can perform well on one measure and poorly on the other. Conceptual framework proposed by the authors; it has not been validated empirically.

Figure 1. Textbook outcome and failure to rescue as complementary measures with different denominators. Textbook outcome is an unconditional all-or-none proportion computed on all eligible patients and is driven by complication avoidance, discharge timing and readmission. Failure to rescue is a conditional proportion computed only among patients who develop a qualifying major complication and reflects recognition and response. The dashed path shows patients who have no major complication but still fail textbook outcome because of prolonged length of stay or readmission; these patients are not in the failure to rescue denominator. The figure is a conceptual framework proposed by the authors and has not been validated empirically. FTR, failure to rescue; LOS, length of stay; TO, textbook outcome.

Paired reporting therefore needs explicit rules, which we propose in Section 7 as an interpretive aid and not as a validated classification. Both measures are computed on the same cohort and period. TO uses all eligible patients. FTR uses the patients with at least

one of the four STS FTR complications (prolonged ventilation, stroke, reoperation and renal failure) as the denominator and the operative deaths among them as the numerator [1]. Each measure is risk-adjusted with its own model, because the predictors of a complication-free course and of death after a complication differ, and each is reported with its denominator and its uncertainty interval. A hospital is described as discordant only when both intervals exclude the reference value in opposite directions, and the FTR denominator at a single center will often be too small for that. In the STS database, 12.0% of patients had at least one of the four complications (derived from the published counts) [1]; for a hospital with 150 eligible operations a year, the annual FTR denominator is therefore about 18 patients, which is why multi-year windows are necessary. Using the same complication definitions and reporting period for both measures has two limitations that should be stated with the results. Any error in complication ascertainment moves both measures at once, so agreement between them does not confirm either. The alignment also restricts FTR to four complications, although deaths after other events, such as mesenteric ischemia, sepsis or cardiac arrest, are also failures of rescue, and restricted FTR definitions omitted 42% to 49% of deaths in a general surgical comparison [44].

4. Textbook Outcome in Adult Cardiac Surgery and in Comparator Fields: The Published Definitions

4.1. Adult Cardiac Surgery

Two registry analyses provide most of the evidence in adult cardiac surgery (Table 1). Verma and colleagues used the Nationwide Readmissions Database (NRD) from 2016 to 2022 and identified an estimated 963,775 elective admissions for CABG and/or valve operations in which the operation took place on the day of admission, after excluding endocarditis, heart transplantation, ventricular assist device implantation and centers with fewer than 52 operations a year [16]. TO was defined, with reference to the STS morbidity endpoints, as survival to discharge without cardiac arrest, stroke, prolonged ventilation above 24 h, acute renal failure, sepsis, pulmonary embolism, reoperation during the index hospitalization or length of stay above 14 days. Every component was ascertained during the index admission from diagnosis and procedure codes. TO was achieved in 86.0% of patients. Prolonged hospitalization was the most common reason for failure and was present in 58.6% of non-TO cases, followed by prolonged ventilation (28.5%) and in-hospital death (13.4%). Only 44.7% of the patients with a prolonged stay had another recorded adverse event, so a stay above 14 days was present in about 8.2% of all patients and was the only failed component in about 4.5% (our calculations from the published percentages). Mortality at 180 days and non-elective readmission were not components. They were outcomes tested for association with TO, and after risk adjustment, TO was associated with a hazard ratio of 0.37 (95% confidence interval [CI] 0.33 to 0.41) for 180-day mortality and 0.66 (0.65 to 0.68) for non-elective readmission. TO also differed by procedure. Relative to isolated CABG, the adjusted odds of TO were higher after isolated valve operations (odds ratio 1.38, 95% CI 1.34 to 1.42) and lower after combined and multivalve operations, and the rates that we calculated from the published table were about 88% after isolated CABG, 88% after isolated valve, 78% after combined CABG and valve and 71% after multivalve operations.

The same study is the only adult cardiac source of hospital-level data. In a multilevel logistic model with patient factors and a hospital-level random effect, the intraclass correlation coefficient (ICC) for TO was 8.5% after adjustment for patient characteristics, compared with 14.5% for in-hospital mortality and for FTR, which was defined as in-hospital death among patients with at least one of the listed complications. Hospital TO rates correlated modestly with in-hospital mortality rates ($r = -0.37$) and minimally with FTR rates

(r = −0.09); the discussion of the paper gives −0.28 and −0.06 for the same correlations. Hospitals in the lowest decile of risk-adjusted TO had observed TO rates of 51.6% to 91.0%, a lower median annual volume (107 versus 137 cases per year) and were less often teaching hospitals (73.6% versus 79.9%). Among their patients who failed TO, prolonged stay was more frequent (69.1% versus 56.9%) and reoperation, renal failure and death were less frequent than at other hospitals. Their patients were more often in the lowest income quartile (36.6% versus 22.8%), although they were slightly younger and had fewer comorbidities. The authors described the center-level variation as substantial and concluded that TO, in-hospital mortality and FTR measure complementary aspects of quality. The discrimination and calibration of the model and the reliability of the hospital-level estimates were not reported. The study supports four statements that recur in this review: TO defined in this way was achieved in 86% of elective patients, length of stay was the most frequent failed component and often the only one, a measurable share of the variation lay between hospitals, and hospitals with low TO rates were distinguished by long stays and not by excess death or major complications.

Ko and colleagues analyzed 3026 primary elective mitral valve operations in the NHR between 2013 and 2021 and built a procedure-specific definition [17]. Mitral TO required the absence of respiratory insufficiency, prolonged intubation, readmission to the intensive care unit, postoperative stroke, renal failure, major vascular complications, mediastinitis and reintervention for bleeding, together with survival and freedom from reintervention at 1 year. TO was achieved in 87.1% of patients (2637 of 3026; calculated 95% CI 85.9% to 88.3%). Factors independently associated with failure were age (odds ratio 1.03, 95% CI 1.01 to 1.03), left ventricular ejection fraction below 50% (1.33, 1.02 to 1.72), pulmonary hypertension (1.30, 1.00 to 1.68), recent myocardial infarction (4.71, 1.79 to 12.38), mitral valve replacement rather than repair (1.89, 1.48 to 2.41), a sternotomy approach (1.30, 1.02 to 1.66) and concomitant rhythm surgery (1.33, 1.03 to 1.71). The definition differs from the American one in two ways. It omits hospital length of stay and hospital readmission, which removes the component that was most frequent among failures in the Verma cohort, and it includes 1-year survival and freedom from reintervention as components, which makes it a 1-year measure and leaves no independent survival outcome against which to test it. The two studies report similar TO rates with component sets that share several major complications but differ in time horizon and in the inclusion of length of stay and readmission. The similarity of the two rates therefore says nothing about the similarity of performance, and this is the first controversy.

The broader search identified four further adult cardiac studies in which TO was an outcome and not the subject of the paper. Two propensity-matched series by the same group compared endoscopic-assisted minimally invasive CABG with a sternotomy approach and defined TO as the absence of 30-day mortality, re-exploration for bleeding, postoperative ischemia, tamponade, cerebrovascular events, wound infection, new-onset arrhythmia, pneumonia, chest drain placement and a hospital stay above 7 days. TO was achieved in 78.1% of 73 endoscopic and 59.1% of 137 sternotomy patients undergoing multivessel total arterial grafting, and in 81.9% versus 59.5% of 116 matched pairs with single-vessel disease, in whom the endoscopic approach was independently associated with TO (odds ratio 3.02, 95% CI 1.61 to 5.66) [18,19]. With a 7-day threshold and new-onset arrhythmia as a component, about 40% of sternotomy patients failed, which shows how much the rate depends on the definition. Two administrative analyses used TO to study disparities. Among 267,914 Medicare beneficiaries undergoing CABG, residence in a county with high food insecurity was associated with lower odds of TO (odds ratio 0.94, 95% CI 0.90 to 0.99) [20]. Among 82,502 surgical or transcatheter aortic valve replacements in the National Inpatient Sample, with TO defined as no in-hospital death, no postoperative complication and no prolonged stay, hospitals in the highest quintile of a racial and

ethnic diversity index had lower odds of TO after surgical replacement (odds ratio 0.79, 95% CI 0.73 to 0.86) [21]. Eleven further claims-based studies pooled CABG with non-cardiac operations and are not specific to cardiac surgery (Table S1).

The gaps in this literature are as informative as its content. No registry-based TO definition has been published for isolated CABG, surgical aortic valve replacement, combined valve and coronary operations or open thoracic aortic surgery, although each has a dedicated STS risk model or a distinct complication profile [8]; the CABG-specific definitions come from two single-group technique comparisons. Both registry analyses are restricted to elective operations. Urgent and emergency cases, in which complications, rescue failures and prolonged stays are concentrated, are therefore unstudied, and the rates, failure drivers and hospital-level variation reported for elective cohorts cannot be extrapolated to them. TO differed by procedure in the national cohort, from about 88% after isolated CABG to 71% after multivalve operations, so a pooled hospital rate depends on procedure mix and cannot distinguish a hospital with excellent CABG results and poor valve results from one with the reverse. The national cohort did not include post-discharge events as components, and the Dutch definition did not include length of stay or hospital readmission. No adult cardiac study has reported the calibration or validation of a risk model for TO, the reliability of hospital-level TO estimates, or a comparison of TO with the STS composite score, and the comparison with FTR is limited to one hospital-level correlation [16]. Every subsequent cardiac TO study will have to make the same definitional choices, and the specialty has not yet agreed on how to make them.

4.2. Comparator Fields: Heart Transplantation and Congenital Heart Surgery

The transplant and congenital series are included for one methodological reason: they show how TO behaves when the definition is strict relative to the expected course. Their rates are not benchmarks for adult cardiac surgery. Zakko and colleagues defined TO after adult heart transplantation in 24,620 United Network for Organ Sharing (UNOS) recipients as a length of stay of 30 days or less, ejection fraction above 50% and functional status of 80% to 100% at 1 year, freedom from rejection, dialysis and stroke during the index admission, and freedom from graft failure, dialysis, rejection, retransplantation and death during the first year [45]. TO was achieved in 45.4% of recipients (95% CI 44.7% to 46.0%) and was associated with a hazard ratio for death of 0.55 (0.50 to 0.59) among 1-year survivors. Bakhtiyar and colleagues applied a different definition, with a stay below 21 days, to 26,885 recipients and found TO in 37%; hospital-specific risk-adjusted TO rates ranged from 39% to 91%, against 97% to 99% for 1-year survival, and 9% of the variation was attributable to differences between hospitals [46]. Prabhu and colleagues built single-center TO definitions for the Norwood and Glenn operations, and Cook and colleagues for the Fontan operation, each requiring length of stay and ventilation duration no longer than the 75th percentile, taken from the STS data report in the Norwood series, together with freedom from operative mortality, reintervention, readmission within 30 days of discharge, extracorporeal support and other major complications during the index hospitalization [22,47,48]. TO was achieved in 30% (58 of 196), 51% (137 of 269) and 49% (97 of 198) of patients. Among patients who failed, prolonged ventilation (49%) and reintubation (46%) were most frequent after the Norwood operation, and prolonged stay (59%) and prolonged ventilation (51%) after the Glenn operation. TO was associated with hazard ratios for long-term death of 0.45 (95% CI 0.22 to 0.89) and 0.50 (0.25 to 0.99) in the Norwood and Glenn series, with a survival difference on log-rank testing in the Fontan series, and with hospital costs that were 43% to 49% lower (total costs in the Norwood series, direct costs in the Glenn and Fontan series). Two points transfer to the adult discussion. A threshold set at a percentile of a reference distribution makes a fixed share of the reference population fail by construction, and a definition that extends to 1 year, as in

transplantation and in the Dutch mitral definition, measures durability as well as perioperative care. In the Zakko definition, 1-year survival is itself a component, so the survival association could be tested only among patients who survived the first year.

4.3. Comparator Fields: Open and Endovascular Aortic Repair

Aortic surgery has the longest registry experience with TO, and its definitions are instructive because they have already confronted the choices that cardiac surgery now faces. Karthaus and colleagues introduced TO for elective abdominal aortic aneurysm repair in the Dutch Surgical Aneurysm Audit (DSAA) in 2017 with procedure-specific length-of-stay limits of 4 days for endovascular repair and 10 days for open repair [11]. TO was achieved in 71% of 4039 endovascular and 53% of 1131 open repairs, and case-mix adjusted TO rates varied from 38% to 89% between hospitals for endovascular repair and from 0% to 97% for open repair. Felsted and colleagues proposed a patient-centered TO for open repair in 9657 Vascular Quality Initiative (VQI) patients that required no in-hospital complication, no reintervention, a postoperative stay of 10 days or less, discharge to home and 1-year survival, and found a TO rate of 44% (4293 of 9657); the five components were achieved by 61.1%, 92%, 78%, 75.8% and 91.2% of patients respectively, so in-hospital complications and not length of stay were the most frequent reason for failure [23]. Sickels and colleagues applied a TO definition to approximately 77,000 elective endovascular, thoracic endovascular and complex endovascular repairs (denominator derived from the published counts) and found TO rates of 77.6%, 69.3% and 61.1%, with a hazard ratio for 1-year death of 0.36 (95% CI 0.33 to 0.39) after standard endovascular repair [24]. Two studies set the length-of-stay component at the 75th percentile of the cohort. In 80,948 elective aneurysm repairs, TO was achieved in 80% of men and 67% of women after endovascular repair and in 54% and 47% after open repair [49]. In 609 fenestrated or branched endovascular repairs, TO was achieved in 64.2%, prolonged stay was the most common failed component, recorded in 23.6%, as a threshold at the 75th percentile of the cohort implies, and 48.2% of failures involved a single component [50]. Lentz and colleagues tested both TO and FTR in elective juxtarenal aneurysm repair across the Dutch and Swedish registries and reported TO of 81.4% after complex endovascular repair and 73.6% after open repair in the Netherlands, with no significant difference in Sweden (88.4% versus 92.5%), where the main structural difference between the registries was the absence of readmission data [25]. A registry that cannot record readmission cannot count it as a failure, so its TO rate is not comparable with that of a registry that can. Figure 2 displays the published rates with the length-of-stay criterion and time horizon of each definition. It is a descriptive display. The cohorts differ in population, procedure, components, thresholds, ascertainment window and data source, and differences between the rates in the figure are differences between definitions and case mix and cannot be read as differences in surgical quality.



Figure 2. Published textbook outcome rates in adult cardiac surgery and in comparator fields, shown with the length-of-stay criterion and time horizon of each definition. Points show the reported proportion of patients achieving textbook outcome and horizontal lines the 95% confidence interval, as published for the Zakko series and otherwise calculated by the Wilson method from published counts; no interval is shown where counts were not reported. Rates are shown to one decimal place and were recalculated from the published counts where these were available. The length-of-stay column gives the stay that counts as a failure. The figure is descriptive: populations, procedures, components, thresholds, ascertainment windows and data sources differ between studies; the rates are not directly comparable; and differences between them do not indicate differences in surgical quality (see Table 1). The sex-stratified aneurysm series and the Swedish juxtarenal series are listed in Table 1 and are not plotted. AAA, abdominal aortic aneurysm; CABG, coronary artery bypass grafting; cEVAR, complex endovascular aortic repair; CI, confidence interval; DSAA, Dutch Surgical Aneurysm Audit; EVAR, endovascular aortic repair; FB-EVAR, fenestrated-branched endovascular aortic repair; LOS, length of stay; NHR, Netherlands Heart Registration; NRD, Nationwide Readmissions Database; OPTN, Organ Procurement and Transplantation Network; P75, 75th percentile; TEVAR, thoracic endovascular aortic repair; TO, textbook outcome; UNOS, United

Network for Organ Sharing; VQI, Vascular Quality Initiative. Data from references [11,16–19,22–25,45–48,50].

Table 1. Published textbook outcome definitions and rates in adult cardiac surgery (primary evidence) and in comparator fields.

Study	Data Source and Cohort	Components of Textbook Outcome	LOS Criterion (Failure If)	Ascertainment Window	TO Rate (95% CI)	Component-Level Failure	Outcomes Tested for Association (Not Components)
Adult cardiac surgery (primary evidence)							
Verma 2026 [16]	NRD 2016 to 2022; 963,775 elective CABG and/or valve admissions (weighted estimate); operation on the day of admission; centers with fewer than 52 operations a year excluded	Survival to discharge; no cardiac arrest, stroke, prolonged ventilation (>24 h), acute renal failure, sepsis, pulmonary embolism or reoperation during the index hospitalization; LOS criterion	>14 days	Index admission	86.0% (828,946/963,775, weighted estimates)	Of failures: prolonged LOS 58.6%, prolonged ventilation 28.5%, in-hospital death 13.4%; 55.3% of prolonged stays had no other recorded event	180-day mortality, HR 0.37 (0.33 to 0.41); non-elective readmission, HR 0.66 (0.65 to 0.68); hospital-level correlation with in-hospital mortality r = −0.37 and with FTR r = −0.09
Ko 2025 [17]	NHR 2013 to 2021; 3026 primary elective mitral operations	No respiratory insufficiency, prolonged intubation, ICU readmission, stroke, renal failure, major vascular complication, mediastinitis or reintervention for bleeding; alive and free of reintervention at 1 year	None	Index course plus 1 year	87.1% (2637/3026; 85.9% to 88.3% *)	Not stated †	None (1-year survival is a component)
Görtzen 2024 [18]	Single center; multivessel total arterial CABG; 73 endoscopic and 137 sternotomy patients after matching	No 30-day mortality, re-exploration for bleeding, postoperative ischemia, tamponade, cerebrovascular event, wound infection, new-onset arrhythmia, pneumonia or chest drain placement; LOS criterion	>7 days	30 days for mortality; other components not stated †	Endoscopic 78.1% (67.3% to 86.0% *); sternotomy 59.1% (50.8% to 67.0% *)	Not stated †	None
Sampon 2024 [19]	Same group; single-vessel CABG; 116 endoscopic and 116 off-pump sternotomy patients after matching	Same definition as Görtzen 2024 [18]	>7 days	30 days for mortality; other components not stated †	Endoscopic 81.9% (73.9% to 87.8% *); sternotomy 59.5% (50.4% to 68.0% *)	Not stated †	None
Heart transplantation (comparator)							
Zakko 2024 [45]	UNOS 2005 to 2017; 24,620 adult heart transplants	EF >50% and functional status 80% to 100% at 1 year; no rejection, dialysis or stroke in the index admission; no graft failure, dialysis, rejection, retransplantation or death in the first year; LOS criterion	>30 days	Index admission plus 1 year	45.4% (11,169/24,620; 44.7% to 46.0%)	Not stated †	Death among 1-year survivors, HR 0.55 (0.50 to 0.59)
Bakhtiyar 2023 [46]	OPTN 2011 to 2022; 26,885 heart transplants	No stroke, pacemaker or dialysis; no ECMO within 72 h; no acute rejection or primary graft dysfunction; no readmission for rejection or infection and no retransplantation within 1 year; EF >50% at 1 year; LOS criterion	≥21 days	Index admission plus 1 year	37% (9841/26,885; 36.0% to 37.2% *)	Not stated †	5-year mortality, HR 0.71 (0.65 to 0.78); hospital range of risk-adjusted TO 39% to 91%
Congenital heart surgery (comparator)							
Prabhu 2023 [47]	Single center 2005 to 2021; 196	No operative mortality, cardiac reintervention, ECMO, cardiac arrest or reintubation	>66 days (75th percentile, STS	Index hospitalization plus 30 days	30% (58/196; 23.6% to 36.3% *)	Of 138 failures: prolonged ventilation 49% (68),	Long-term death, HR 0.45 (0.22 to

	Norwood operations	before discharge; no readmission within 30 days of discharge; LOS and ventilation criteria	data report); ventilation >10 days	after discharge		reintubation 46% (63), prolonged LOS 30% (41); one component failed in 34%, two in 26%	0.89); total hospital cost 43% lower
Prabhu 2024 [48]	Single center 2005 to 2021; 269 Glenn operations	No operative mortality, reintervention, 30-day readmission, ECMO or major thrombotic complication; LOS and ventilation criteria	>17 days (75th percentile); ventilation >2 days	Index hospitalization plus 30-day readmission	51% (137/269; 45.0% to 56.8% *)	Of 132 failures: prolonged LOS 59% (78), prolonged ventilation 51% (67)	Long-term death, HR 0.50 (0.25 to 0.99); direct cost 48% lower
Cook 2025 [22]	Single center 2005 to 2022; 198 Fontan operations	No operative mortality, reintervention, 30-day readmission, ECMO, major thrombotic complication or chylothorax; LOS and ventilation criteria	>75th percentile (value not stated †)	Index hospitalization plus 30-day readmission	49% (97/198; 42.1% to 55.9% *)	Not stated †	Long-term survival, log-rank p = 0.027; direct cost 49% lower
Open and endovascular aortic repair (comparator)							
Karthaus 2017 [11]	DSAA 2014 to 2015; 4039 EVAR and 1131 open elective AAA repairs	No intraoperative complication, postoperative surgical complication, reintervention, readmission or mortality; LOS criterion	>4 days EVAR; >10 days open	Mortality within 30 days of surgery or at discharge; readmission window not stated †	71% EVAR; 53% open (counts not stated †)	Qualitative: prolonged LOS, postoperative complications, readmission	None; adjusted hospital range 38% to 89% (EVAR) and 0% to 97% (open)
Felsted 2024 [23]	VQI 2009 to 2022; 9657 elective open AAA repairs in 198 centers	No in-hospital complication; no reintervention or reoperation; discharge to home; 1-year survival; LOS criterion (readmission is not a component)	>10 days	Index admission plus 1 year	44% (4293/9657; 43.5% to 45.4% *)	Share of all patients failing: complication 38.9%, reintervention 8%, LOS 22%, non-home discharge 24.2%, death within 1 year 8.8%	None independent (1-year survival is a component)
Sickels 2026 [24]	VQI; elective EVAR 2003 to 2023, TEVAR and cEVAR 2011 to 2023; about 77,000 procedures (derived)	No major complication, perioperative reintervention or mortality; discharge to previous residence; LOS criterion	Prolonged LOS (threshold not stated †)	Perioperative	77.6% EVAR; 69.3% TEVAR; 61.1% cEVAR	Not stated †	1-year death, HR 0.36 (0.33 to 0.39) for EVAR and 0.24 (0.20 to 0.30) for cEVAR
Filiberto 2025 [49]	VQI; 80,948 elective AAA repairs	No major medical or surgical complication, reintervention, mortality or non-home discharge; LOS criterion	>75th percentile of the cohort	Not stated †	EVAR: men 80%, women 67%; open: men 54%, women 47%	Qualitative (women): prolonged LOS, surgical complications, non-home discharge	1-year survival
Oden-Brunson 2026 [50]	Prospective device study 2013 to 2025; 609 fenestrated-branched endovascular repairs	No major complication, early reintervention, non-home discharge or early mortality; LOS criterion	>75th percentile of the cohort	30 days or index hospitalization	64.2% (391/609; 60.3% to 67.9% *)	Prolonged LOS 23.6%; major complications 17.9%; single-component failure in 48.2% of failures	Overall survival, adjusted HR 0.58 (0.43 to 0.78)
Lentz 2026 [25]	DSAA (n = 1925) and Swedvasc (n = 775) 2016 to 2023; elective juxtarenal AAA repair	Ideal postoperative course (component list not stated †); FTR defined as death after a Clavien-Dindo grade IV or higher complication	Not stated †	Not stated †; readmission not recorded in Swedvasc	NL: 81.4% cEVAR, 73.6% open; SE: 88.4% cEVAR, 92.5% open	Not stated †	None; FTR 1.6% cEVAR and 4.3% open (NL)

Components are listed as defined by each study. The length-of-stay criterion is a component in every study that has one and is shown in its own column as the stay that counts as a failure. Outcomes in the last column were tested for association with TO and are not components. Two adult cardiac claims-based studies that used TO as an outcome without reporting a TO rate are described in the

text [20,21]. * 95% CI calculated by the authors with the Wilson method from published counts; for the two coronary series the numerators were derived from the published percentages. † As given in the published abstract. AAA, abdominal aortic aneurysm; CABG, coronary artery bypass grafting; cEVAR, complex endovascular aortic repair; CI, confidence interval; DSAA, Dutch Surgical Aneurysm Audit; ECMO, extracorporeal membrane oxygenation; EF, ejection fraction; EVAR, endovascular aortic repair; FTR, failure to rescue; HR, hazard ratio; ICU, intensive care unit; LOS, length of stay; NHR, Netherlands Heart Registration; NL, Netherlands; NRD, Nationwide Readmissions Database; OPTN, Organ Procurement and Transplantation Network; SE, Sweden; TEVAR, thoracic endovascular aortic repair; TO, textbook outcome; UNOS, United Network for Organ Sharing; VQI, Vascular Quality Initiative.

5. Controversy 1: The Definition Problem

No two cardiac or aortic TO definitions in Table 1 are the same, and the differences are not cosmetic. Length-of-stay thresholds range from 4 days to 66 days, are set at a percentile, or are omitted. Readmission is counted within 30 days of discharge, within an unstated window, or not at all. Mortality is in-hospital, operative, 30-day or 1-year. Complications are a fixed registry list or any complication, and in the oncology literature often a Clavien-Dindo grade. The same problem has been documented in surgical oncology, where a systematic review of 86 papers covering 116 disease-specific studies found that no component was included in every study and that mortality, complications, readmission and length of stay appeared in 88.8%, 90.5%, 93.1% and 89.7% of definitions with widely varying thresholds [51]. A meta-analysis in cancer surgery found an association between TO and survival for most tumor sites, with exceptions in pancreatic, gallbladder and ampullary cancer [52], and TO rates vary with the definition: a mean of 49.4% (range 21.8% to 84.2%) in esophagogastric surgery and medians of 19.5% to 62.0% across hepatopancreatobiliary procedure groups [53,54]. The gastrectomy community has responded with an international Delphi consensus to fix the parameters, and distal pancreatectomy has been defined by systematic review followed by a clinical relevance check [55,56]. Cardiac surgery has not done either.

The component labels also conceal differences in operational definition, and the same label does not mean the same event across registries or across versions of one registry (Table S2). Operative mortality in the STS database is death during the index hospitalization at any time, including after transfer to another acute care facility, or death after discharge within 30 days of the operation, and in the current data version it includes discharge to hospice [27]. The EACTS Adult Cardiac Database records death within 30 days of surgery irrespective of location and derives in-hospital death from the status at discharge [28]. Permanent stroke was a deficit lasting more than 72 h in the earlier STS risk models and is now a deficit that does not resolve within 24 h or is confirmed on imaging [9,27]. Prolonged ventilation is more than 24 h in the STS database and more than 48 h in the EACTS specification. Renal failure in the STS database is a threefold rise in creatinine, a creatinine of 4.0 mg/dL or more with a rise of at least 0.5 mg/dL, or new dialysis, whereas the EACTS specification records new dialysis. Reoperation in the STS major morbidity endpoint changed in the 2018 models from reoperation for any reason to reoperation for bleeding, tamponade or any cardiac reason, and it excludes planned staged procedures and non-cardiac returns to the operating room [9,27]. In this review, mortality, stroke, renal failure, prolonged ventilation, reoperation, deep sternal wound infection, prolonged length of stay and readmission are used in the current STS sense unless a study's own

Three design decisions drive most of the variation. The first is the length-of-stay criterion, where the choice is between an absolute threshold and a percentile of a reference distribution. An absolute threshold has a fixed meaning over time and across hospitals. The STS risk models and the EACTS specification both define a prolonged stay as more

than 14 days from the operation to discharge [9,28]. In the STS model development sample, this event occurred in 5.0% of CABG, 8.0% of valve and 12.9% of combined valve and CABG operations [9], and in the national TO cohort, a stay above 14 days was present in about 8.2% of patients and was the only failed component in about 4.5% [16]. A percentile threshold behaves differently, and the arithmetic deserves to be explicit. If the threshold is the p-th percentile of the population to which it is applied, about (100 − p)% of patients fail that component by construction, so TO cannot exceed about p% whatever the quality of care. A 75th percentile threshold caps TO near 75% before any complication is counted, as the two aortic series that used one illustrate [49,50], and a rate obtained in this way cannot be compared with the 86% to 87% obtained with a 14-day threshold or with none. A percentile also redefines failure whenever the reference distribution is updated, so a hospital's TO rate can change with no change in its care, and it ranks hospitals against each other instead of against a standard. A percentile is defensible only when it is derived once from an external reference population and period and then frozen as a number of days, at which point it has become an absolute threshold. Procedure-specific absolute thresholds, such as the 4 and 10 days used for endovascular and open aneurysm repair [11], follow the same logic. For these reasons, the core set proposed in Section 10 uses the existing 14-day definition, and percentile-based or shorter procedure-specific thresholds are reserved for prespecified sensitivity analyses.

Whatever threshold is chosen, length of stay is the component most exposed to factors other than the quality of the operation, and each of them needs an explicit rule. Case mix and urgency: a single threshold is failed more often after combined, redo and urgent operations, which is a reason to report by procedure group and admission status and to adjust for patient risk, and not a reason to set a more lenient threshold for each subgroup, because that would hide the difference it is meant to describe. Minimally invasive surgery: a 14-day threshold cannot register the shorter recovery that such approaches aim for, whereas the 7-day threshold used in the endoscopic CABG series can [18,19]; this is an argument for reporting a continuous measure beside the binary one. Transfers: the STS definition follows a transferred patient to discharge from the receiving acute care facility, so the stay is not shortened by transfer [27], but registries that close the record at transfer will undercount long stays, and health systems that routinely return patients to a referring hospital are affected most. Rehabilitation pathways and discharge practice: hospitals with early access to inpatient rehabilitation or skilled nursing discharge sooner, and the length of the acute stay is partly a property of the local health system. The same patient can therefore pass TO in one system and fail in another. We propose that the stay be counted from the operation to discharge from acute care, that discharge destination and days at home be reported beside TO, and that every hospital comparison be repeated with the length-of-stay component removed.

The second decision is the time horizon, and with it the consistency of the ascertainment windows. A core set built from registry fields inherits three windows: the index hospitalization for stroke, renal failure, prolonged ventilation and reoperation; the index hospitalization or 30 days from surgery for mortality and deep sternal wound infection; and 30 days from discharge for readmission. The mixture has a rationale, since each window is the one in which the registry already ascertains the event and re-abstraction to a common window is not feasible at scale, but it has consequences that a report should state. Patients are observed for different lengths of time: a patient discharged on day 5 is followed to day 35 and a patient discharged on day 40 to day 70, and complications after discharge are captured only if they cause readmission or death. Patients with a prolonged stay have already failed TO, so their later events do not change their TO status but do change the component profile. A single window of 30 or 90 days from surgery for all components would be cleaner and is the structure used by days at home, and a core set should

state its windows component by component, as the proposal in Section 10 does. Survival and reintervention at 1 year belong in the assessment of a valve operation, whose purpose is durable valve function, as Ko and colleagues recognized [17]. They should not sit inside a perioperative TO, where they would mix surgical quality with disease progression and with the completeness of follow-up, and we treat them as separately reported durability indicators.

The third decision is readmission, and specifically its time anchor. Readmission within 30 days of discharge and readmission within 30 days of surgery are different events. The STS readmission field was defined from the date of surgery in earlier data versions and has been defined from the date of discharge since version 2.81 in July 2014; it counts admission to any acute care hospital, planned or unplanned, and excludes observation stays, rehabilitation and skilled nursing [27]. The STS readmission measure for CABG takes the outcome from Medicare claims linked to the registry, and its development study reported a median hospital risk-standardized rate of 16.8% (range 12.6% to 23.6%) among Medicare beneficiaries [57]. A discharge anchor gives every survivor the same 30 days at risk but shifts the window later for patients with long stays; a surgery anchor gives a common calendar window but leaves a patient who stays 25 days only 5 days at risk. Rates from the two anchors are not interchangeable, and the anchor of a published TO definition is often not stated (Table 1). We propose the discharge anchor for the core set because it matches the current STS field and the claims-based measure, and a surgery-anchored 30-day window for days at home. Hannan showed in 2003 that 12.9% of New York CABG patients were readmitted within 30 days, most often for infection and heart failure, and that readmission was related to surgeon volume and hospital mortality after adjustment [58]. Readmission is a quality signal and a patient-relevant event, and the Dutch and Swedish comparison shows what is lost when a registry cannot record it [25].

6. Controversy 2: High Achievement Rates, Discrimination and Risk Adjustment

TO was achieved in 86% to 87% of elective patients in the two registry analyses [16,17]. A high achievement rate raises the concern of a ceiling effect, and the concern needs to be stated precisely, because a high prevalence does not by itself establish poor discrimination between hospitals. Whether a hospital-level measure discriminates depends on four quantities: the variance of true performance between hospitals, the number of cases per hospital, the prevalence of the outcome, which sets the binomial sampling variance, and the precision of the risk adjustment. Prevalence acts in both directions. The sampling variance of a proportion, $p(1 - p)/n$, shrinks as p approaches 1, but so does the room for hospitals to differ, and the balance between the two is an empirical question. There is no threshold of prevalence above which a measure stops discriminating, and no empirical basis for a cutoff at 80% or at any other value. What theory supports is limited to two statements: an all-or-none measure near 100% compresses the absolute differences between hospitals, and the ICC of a binary outcome on the observed scale falls as its prevalence moves away from 50% [59,60]. The fact that this limits the usefulness of TO in adult cardiac surgery is a theoretical concern and has not been demonstrated.

The available evidence does not indicate that a ceiling prevents discrimination. In the national cohort, the ICC for TO was 8.5% after adjustment for patient factors, hospitals in the lowest decile had observed TO rates of 51.6% to 91.0% and differed in volume and teaching status, and the authors concluded that center-level variation was substantial [16]. The ICC was larger for in-hospital mortality and for FTR (14.5%), so on that scale TO did not separate hospitals more than mortality did. In heart transplantation, hospital-specific risk-adjusted TO rates ranged from 39% to 91% while 1-year survival ranged from 97% to 99% [46]. Neither study reported the reliability of the hospital-level estimates, which is

the quantity that determines whether observed differences reflect performance or chance. Reliability is the ratio of the variance between hospitals to the total variance of the observed hospital rates. It runs from 0 to 1, and 0.70 is commonly taken as the minimum for profiling [61]. It falls with small caseloads and low event rates: in an analysis of six general and vascular operations, 37.7% of hospitals reached 0.70 for morbidity after colon resection and 7.1% after pancreatic resection [62]. An illustrative calculation shows the order of magnitude for TO (Supplementary Note S1). If the ICC of 8.5% is on the latent logistic scale, which the paper does not state but which is the usual output of multilevel logistic models, it corresponds to a median odds ratio of about 1.7 and, by linearization, to an observed-scale ICC of about 0.036 at a prevalence of 86% [63,64]. Under those assumptions, the reliability of a hospital's TO rate would be about 0.65 with 50 cases, 0.79 with 100 cases and 0.83 with 137 cases, and about 64 cases would be needed to reach 0.70. The same calculation for in-hospital mortality, with an ICC of 14.5% and a prevalence of about 1.9%, gives a reliability of about 0.58 with 137 cases, because a rare event carries less information per patient. These figures are an illustration under stated assumptions and not an estimate from data. They suggest that the high prevalence of TO need not preclude reliable hospital-level measurement at ordinary annual volumes, and they show that reliability has to be estimated and reported and cannot be assumed in either direction.

We therefore propose an explicit analytic framework for hospital comparison, drawn from the methods already used for publicly reported outcome measures [65,66]. The patient-level model is a hierarchical logistic regression of TO on patient risk factors with a random intercept for hospital, fitted separately for each procedure group over a rolling 3-year window. The hospital result is the risk-standardized TO rate, calculated as the ratio of the number of TO cases predicted with the hospital's own intercept to the number expected with the average intercept for the same patients, multiplied by the overall TO rate. This estimate is shrunk toward the overall rate in proportion to the hospital's imprecision, which is what reliability adjustment means [3,66]. Three quantities are reported with it. The first is between-hospital variation, as the variance of the random intercept, the variance partition coefficient with the scale on which it was calculated, and the median odds ratio [63,64]. The second is the reliability of each hospital's estimate, with a prespecified minimum (we suggest 0.70) or a minimum caseload below which a hospital's rate is shown but not classified [61,62]. The third is uncertainty: a 95% interval for each hospital's risk-standardized rate, obtained as a Bayesian credible interval or by bootstrap resampling, with a hospital classified as different from the reference only when its interval excludes the overall rate, which is the rule used for STS star ratings [7,67,68]. Funnel plots of hospital rates against caseload with control limits display the same information without a league table [69]. Rank lists and comparisons of point estimates are avoided. Shahian and colleagues showed that four common methods for calculating hospital-wide mortality classified the same hospitals differently: in 2006, 12 of the 28 hospitals with higher-than-expected mortality by one method had lower-than-expected mortality by one or more of the others [70]. TO will inherit this sensitivity to method, and the method should be fixed before the data are examined.

Risk adjustment is the part of this framework that can least be borrowed. The STS predictor sets are validated for operative mortality and for the individual STS endpoints, and EuroSCORE II for mortality [2,9]; neither was built for a composite in which prolonged stay and readmission are frequent components. The national TO analysis adjusted with a multilevel model containing age, sex, comorbidities and operation type and did not report its discrimination or calibration [16], models for TO in two comparator fields reported c-statistics of 0.70 and 0.74 [23,47], and no adult cardiac study has reported the calibration or validation of a risk model for TO. The predictors of a prolonged stay or a readmission are not those of death. In a meta-analysis of 62 studies in 16,679 patients,

frailty predicted discharge to a care facility (odds ratio 5.52) more strongly than short-term mortality (2.33) or readmission (2.00) [71], and the STS readmission model, built on registry data, reached a c-statistic of only 0.648 [57]. A TO model that reuses mortality predictors is therefore likely to under-adjust for the components that fail most often. A TO-specific model should be developed and reported as a prediction model study under the TRIPOD+AI statement [72]. Development requires a prespecified outcome definition and time origin, candidate predictors chosen from the STS or EuroSCORE II variables with the addition of frailty, functional status and preoperative living situation, a sample size planned for the number of candidate parameters, and transparent handling of missing data [73,74]. Performance is assessed by calibration first, as calibration-in-the-large, calibration slope and a calibration plot across the range of predicted risk, and then by discrimination with the c-statistic, with internal validation by bootstrapping, temporal validation in later years and external validation in a second registry [74]. Calibration within procedure groups and within urgent and emergency cases matters more than a single overall c-statistic, because a model that is miscalibrated in high-risk patients will penalize the hospitals that accept them. Two designs should be compared: a single model for the composite, and separate models for each component whose predictions are combined, since the second can use the existing STS component models. Until such a model exists, risk-adjusted TO comparisons between hospitals rest on models whose calibration for TO is unknown.

A worked example shows how sensitive the composite is to a single frequent component. Take the national failure profile, in which 14 of every 100 patients fail TO, about 8 of them with a prolonged stay and between 4 and 5 with a prolonged stay as their only failed component [16]. If a discharge pathway moves those patients below the threshold, the hospital's TO rate rises from 86% to about 90% without any change in complications or deaths. A second hospital that halves its stroke rate from 2% to 1%, which is a large clinical achievement, raises its TO rate by at most 1 point. The payoff from improving a component scales with how common that component is among failures. In the administrative definition, the frequent component was length of stay, and hospitals in the lowest decile of TO were distinguished by prolonged stays and not by more deaths or reoperations [16]. In a definition built on STS fields, the balance is likely to differ, because prolonged ventilation above 24 h alone occurred in 9.3% of CABG and 18.9% of combined valve and CABG operations in the STS development sample, against 5.0% and 12.9% for a stay above 14 days [9]. In either case, a hospital's TO rate can be moved by its discharge practice, its rehabilitation capacity or its extubation protocol with no change in death or major morbidity. This is not a reason to remove length of stay, which patients care about and which carries real cost. It is the reason that the component profile must accompany every TO rate.

The all-or-none rule raises four further issues that can be addressed analytically (Section 3.1). Dominance by the most frequent component is examined by reporting, for every TO rate, the proportion of patients failing each component, the proportion failing each component alone, and TO recalculated with each component removed in turn; in the national cardiac cohort 55.3% of patients with a prolonged stay had no other recorded adverse event, in fenestrated endovascular repair 48.2% of failures involved a single component, and in open aneurysm repair 1574 of 5364 failures were due to an in-hospital complication alone [16,23,50]. Correlation between components is described by the cross-tabulation of failures and by the distribution of the number of failed components per patient, which separates a hospital with many single failures from one with fewer but multiple failures. Unequal clinical importance cannot be captured within a binary measure. It is handled by reporting mortality and major morbidity beside TO and by a sensitivity analysis on an ordinal scale, for example a desirability of outcome ranking (DOOR) with levels

for TO, failure on stay or readmission only, one major complication, several major complications and death [75]. Finally, hospital comparisons should be repeated with a continuous outcome, days at home within 30 days of surgery (DAH30), which uses the information in length of stay, readmission and death without a threshold [76]. If hospital classifications agree across the binary, ordinal and continuous analyses, the binary measure can be used for its simplicity. If they disagree, the disagreement is itself a finding and should be reported.

7. Controversy 3: Textbook Outcome Against the Metrics Already in Use

Cardiac surgery already has a dense measurement infrastructure, and a new metric must show what it adds (Table 2). The STS composite scores combine risk-adjusted operative mortality with risk-adjusted major morbidity and, for CABG, use of the internal mammary artery and perioperative medications, and classify participants with one, two or three stars according to whether the 95% credible interval of the composite lies below, across or above the STS average, using 3 years of data [4,5,7,67]. The morbidity domain is itself scored any-or-none at patient level: a patient counts as free of major morbidity only with none of reoperation, stroke, renal failure, deep sternal wound infection and prolonged ventilation [5,7]. The overlap with TO is therefore large. A core TO built on STS fields differs from the mortality and morbidity domains of the STS composite in three respects only. It adds prolonged stay and readmission, neither of which is part of an STS composite score [27]. It joins death and morbidity in one patient-level event, where the composite weights the two domains. It yields a proportion of patients, where the composite yields a score and a star category. What TO could identify that the existing measures do not is, accordingly, specific: hospitals whose patients avoid death and major morbidity but recover slowly or return to hospital, and a single figure that patients and administrators can read without statistical training. Both are plausible and neither has been demonstrated. No study has compared TO with the STS composite, and the incremental value of TO is at present a hypothesis. The only hospital-level comparison available is with in-hospital mortality and FTR in the administrative cohort, where the correlations with TO were −0.37 and −0.09 [16]. This shows that TO orders hospitals differently from mortality. It does not show which ordering is closer to quality. The evidence needed to test it is a head-to-head analysis in one registry that reports the correlation between hospital TO rates and composite scores, the proportion of hospitals reclassified, the reliability of each measure at realistic caseloads, the ability of each to predict the hospital's results in the following period and its patients' 1-year survival and days at home, and whether feedback of TO changes practice. The STS composite sets a demanding standard, since its CABG version reaches a reliability of 0.78 [67]. If TO adds nothing on these criteria, the case for it rests on communication alone.

FTR is the second established comparator. The STS FTR metric is risk-adjusted, uses four complications, and identified 5.6% of participants as performing worse and 4.7% as performing better than expected [1]. The pairing of TO and FTR as quality indicators has been evaluated explicitly in one registry study, in juxtarenal aortic repair, where the authors found it feasible and concluded that the added value is conceptual and requires further validation [25]; other aortic and claims-based studies have reported both measures as separate endpoints (Table S1). In adult cardiac surgery, the national cohort reported both measures, found minimal correlation between hospital TO and FTR rates, and concluded that the two are complementary and not redundant [16]. We take the same position. The argument for pairing is that the two measures point to different parts of the care pathway. In Michigan, low-mortality hospitals differed from high-mortality hospitals in rescue and not in the incidence of complications, and the collaborative model of structured review

and feedback has produced measured gains in processes and outcomes [38,77]. Prolonged stay and readmission, the components that distinguish TO from the STS morbidity domain, are outcomes that enhanced recovery programs aim to improve [78]. That a quality program reporting TO and FTR together would identify the right lever more often than a program reporting the STS composite with its component rates is an expectation, and it has not been tested. The rules for joint calculation and the reading of discordant results are set out in Section 3.2 and Table 3.

Ordinal and continuous measures are the third alternative. DOOR places each patient on an ordered scale from the best to the worst outcome and compares distributions between groups, and the win ratio compares pairs of patients on a hierarchy of clinical priorities [75,79]. Both preserve information that a binary TO discards, namely the difference between a patient who fails TO on a 15-day stay and one who fails on a stroke and dialysis. Both are harder to communicate, and neither has an accepted cardiac surgical hierarchy. Days alive and at home is a continuous alternative, and because published definitions vary, it must be defined precisely. A scoping review of 78 perioperative studies found 12 different terms for the measure; 40.5% of studies subtracted days in hospital and in other institutions, 36.7% subtracted hospital days only, and deaths were handled in at least four ways [80]. We use the definition validated by Myles and colleagues for DAH30, and we treat days alive and out of hospital (DAOH) as a synonym only when it is calculated in the same way [76]. The day of surgery is day 0. DAH30 equals 30 minus the days spent in the index hospitalization after surgery, minus the days of any readmission to an acute care hospital within 30 days of surgery, counted up to day 30. A transfer to another acute care hospital is a continuation of hospital days. A patient who dies within 30 days of surgery, in or out of hospital, is assigned 0. Days in inpatient rehabilitation, a skilled nursing facility or a long-term acute care hospital are not days at home and are subtracted when the data source records them; when it does not, the measure counts days out of acute hospital only and the report must say so, as the validation study itself had to do [76]. In that study, the median DAH30 after cardiac surgery was 22.8 days. In a Swedish cohort of 636,885 surgical patients, 8 or fewer days at home carried a hazard ratio of 6.78 for death within 1 year compared with 29 or 30 days, and in an Australian cardiac center the median DAOH at 30 days was 21 days and 10.6% of patients spent no time at home [81,82]. DAH30 carries the length-of-stay and readmission information of TO in one continuous number, at the cost of a skewed distribution with a cluster at 0, and we place it beside TO as a supplementary indicator and as a sensitivity outcome, not inside the composite.

Table 2. Textbook outcome compared with the quality metrics already in use in adult cardiac surgery.

Attribute	Textbook Outcome	Failure to Rescue	STS Composite Score	DOOR and Win Ratio	Days at Home Within 30 Days (DAH30)
Question answered	What proportion of patients had the ideal course?	What proportion of patients with a major complication died?	How do risk-adjusted mortality and morbidity compare with the STS average?	Which group has the more desirable distribution of outcomes?	How many of the first 30 days after surgery did the patient spend alive and at home?
Denominator	All eligible patients	Patients with a qualifying major complication	All eligible patients	All eligible patients, ranked or paired	All eligible patients
Scoring	Binary, all-or-none	Binary, conditional	Weighted composite of risk-adjusted domains; morbidity domain scored any-or-none	Ordinal hierarchy	Continuous, 0 to 30 days
Direction	Positive (success)	Negative (failure)	Higher score is better	Ordered	Positive

Most frequent determinant	Prolonged LOS in the one cohort that reported it [16]; unknown for registry-based definitions	Death after prolonged ventilation, stroke, reoperation or renal failure [1]	Operative mortality and major morbidity [5]	Depends on the hierarchy	LOS, readmission and death
Risk adjustment	Multilevel administrative model in one cohort, performance not reported [16]; no validated TO-specific model	STS FTR model, c-statistic 0.806 [1]	STS procedure-specific models [9]	Covariate adjustment possible; no cardiac surgical model	Regression methods for bounded, skewed data
Published distribution	86.0% and 87.1% in two elective registry cohorts; 59.1% to 81.9% in two coronary series [16–19]	6.6% to 13.9% across hospital mortality groups [38,40]	CABG composite: 11.4% of participants one-star and 20% three-star with 3-year data and 95% CrI [67]	Not reported	Median 21 to 22.8 days after cardiac surgery [76,82]
Hospital-level variation and reliability	ICC 8.5% in one cohort [16]; reliability not reported	ICC 14.5% in one cohort [16]; reliability not reported in cardiac surgery, 0.32 in general surgery [44]	0.78 for the CABG composite [67]	Not reported	Not reported
Data requirements	Registry complication fields, LOS and readmission; readmission needs complete abstraction or linkage	Registry complication fields and mortality	Registry fields	Registry fields plus an agreed hierarchy	Dates of surgery, discharge, readmission and death; institution days if counted
Interpretability for patients	Single proportion; summary format preferred in one patient survey [83]	Requires explanation of the conditional denominator	Star category; underlying score not intuitive	Requires explanation	Intuitive unit (days)
Sensitivity to social risk	Shown in claims data [20,21,84]; expected through LOS and readmission	Not established	Not established	Not established	Shown for DAOH after CABG [85]
Status in adult cardiac surgery	Two registry studies and four smaller or claims-based studies; no consensus definition	Risk-adjusted metric developed by the STS Quality Measurement Task Force [1]	Established; voluntary public reporting [86]	No agreed cardiac hierarchy [75,79]	Validated in mixed surgical cohorts; single-center cardiac data [76,82]

Entries describe published evidence in adult cardiac surgery unless stated otherwise; "not reported" means that we found no published estimate. CABG, coronary artery bypass grafting; CrI, credible interval; DAH30, days at home within 30 days of surgery; DAOH, days alive and out of hospital; DOOR, desirability of outcome ranking; FTR, failure to rescue; ICC, intraclass correlation coefficient; LOS, length of stay; STS, Society of Thoracic Surgeons; TO, textbook outcome.

Table 3. Proposed joint interpretation of risk-adjusted textbook outcome and failure to rescue computed on the same cohort and period.

Pattern Relative to the Reference	Possible Interpretation	Alternative Explanations to Exclude	Data to Examine First
TO higher, FTR lower	Complication avoidance, recovery pathway and rescue are all favorable	Low-risk case selection not captured by the models; under-recorded complications or readmissions	Case mix and urgency; completeness of complication and readmission capture
TO higher, FTR higher	Few patients deviate from the ideal course, but those with a	Small FTR denominator; a complicated group made sicker by successful prevention of minor complications;	Number of complicated patients; mortality by complication; timing

	major complication die more often	early discharge that moves events to the period after discharge	of deaths; readmissions and deaths after discharge
TO lower, FTR lower	Rescue is effective; failures arise from the incidence of complications, length of stay or readmission	Longer stays caused by rehabilitation access, transfer practice or social risk and not by complications; more complete recording of complications, which enlarges the FTR denominator	Share of TO failures due to LOS or readmission alone; discharge destination; component rates
TO lower, FTR higher	Both avoidance and rescue need review	High-risk referral case mix beyond the adjustment models; data quality	Risk profile and model calibration in the risk range of the hospital; structural review of intensive care and escalation
Interval of either measure includes the reference	No evidence of a difference	Insufficient volume	Extend the reporting window; do not classify

The table is an interpretive aid proposed by the authors. It has not been validated, and the patterns have not been shown to be stable over time or to correspond to differences in structure or process. "Higher" and "lower" refer to a 95% interval that excludes the reference value. FTR, failure to rescue; LOS, length of stay; TO, textbook outcome.

8. Controversy 4: Patient-Centeredness and Equity

TO is often described as patient-centered, and the description needs qualification. The components reflect what clinicians assume patients want: to survive, to avoid complications, to go home on time and to stay home [10]. The direct evidence of patient preference is thin. In a Dutch survey of 1329 surgical patients, 52.1% preferred a summary measure expressed as the percentage of patients with a textbook outcome and 38.0% preferred separate detailed measures, which supports the format but says nothing about which components patients would choose [83]. No published cardiac TO definition reports patient or public involvement in the choice of components, and neither does the framework proposed here, whose components were chosen by the authors from registry fields. Patient priorities should enter through the consensus process, with patients and caregivers as a stakeholder group in a Delphi exercise conducted to the standards for core outcome set development and with their involvement reported according to the GRIPP2 checklist [87–89]. A second gap is that no cardiac TO definition includes a patient-reported outcome. Discharge to home appears in the aortic definitions as a proxy for functional recovery [23,24], functional status at 1 year appears in a transplant definition [45], and patient-reported measures have been systematically reviewed in congenital cardiac surgery, where only six studies qualified [90]. Validated instruments exist and are specified in international standard sets. For coronary artery disease, the International Consortium for Health Outcomes Measurement (ICHOM) set uses the 7-item Seattle Angina Questionnaire, the Rose Dyspnea Score and the 2-item Patient Health Questionnaire [91,92]. For heart valve disease, the ICHOM set uses the EQ-5D-5L and the Heart Valve Disease Impact on Daily Life questionnaire, collected at baseline, at 3 months and annually [93,94]. The 12-item Kansas City Cardiomyopathy Questionnaire is a validated option for patients with heart failure symptoms [95]. Integration of such measures into the STS database has been discussed as a future step and is not routine [96]. We therefore list a patient-reported measure as a supplementary indicator, reported as the change from baseline at 3 months, and not as a component of the composite, because response rates and timing would otherwise determine the TO rate.

The second issue is equity. Socioeconomic position predicts risk-adjusted death after cardiac surgery independently of race, as Koch and colleagues showed in 23,330 Cleveland Clinic patients, and the Distressed Communities Index predicts risk-adjusted mortality after isolated CABG in Virginia independently of STS predicted risk [97,98]. Social

context predicts TO itself. Among 852,128 Medicare beneficiaries undergoing aneurysm repair, CABG, colectomy or lung resection, a 10% increase in college-educated residents was associated with 4% higher odds of TO and a 10% increase in single-parent households with 2% lower odds, with the strongest effects on extended stay and 90-day mortality [84]. In the national cardiac cohort, patients in lower income quartiles and patients insured by Medicare or Medicaid had lower adjusted odds of TO, and hospitals in the lowest decile of TO treated more patients from the lowest income quartile (36.6% versus 22.8%) [16]. Among Medicare beneficiaries undergoing CABG, high county-level food insecurity was associated with lower odds of TO and with longer stays [20], and hospitals serving a higher proportion of non-white patients had lower odds of TO after aortic valve replacement [21]. DAOH after CABG in New Zealand showed an equity gap between Māori and non-Māori patients that widened from 30 to 365 days and persisted after multivariable adjustment [85]. Prolonged stay and readmission, the components that most often cause TO failure in administrative data, are the components most sensitive to social disadvantage and to the organization of care after discharge.

Two different questions are at stake, and they call for different analyses that should be reported together. The first is the fair comparison of hospitals. If social risk raises the probability of a prolonged stay or a readmission through mechanisms outside the hospital's control, a comparison that ignores it penalizes hospitals that serve disadvantaged populations and invites the risk aversion that public reporting has been associated with in interventional cardiology and cardiac surgery [99–101]. In the Medicare readmission program, adding social risk factors to the models halved the difference between safety-net and other hospitals and reduced the penalties of more than half of safety-net hospitals [102]. The second question is whether disadvantaged patients receive worse care or have worse outcomes. Adjustment answers the first question by removing exactly the differences that the second question asks about. As the developers of the readmission measures argued, adjusting for socioeconomic status could obscure disparities and credit a hospital that delivers inferior care to poor patients with caring for a higher-risk population [103]. Neither analysis can replace the other, and we propose dual reporting. The primary risk-standardized TO rate is adjusted for clinical risk only and is shown overall and stratified by a social risk indicator, such as an area deprivation index, insurance type or dual eligibility, within each hospital and across hospitals, so that disparities remain visible. A second, social-risk-adjusted rate is calculated for any use that carries a financial or reputational consequence; alternatively, hospitals are compared within peer groups defined by their proportion of socially at-risk patients. The within-hospital difference in TO between socially at-risk and other patients is reported as an equity indicator in its own right. Whether TO is more sensitive to social risk than the STS composite has not been tested and is part of the research agenda.

Fair comparison across health systems raises a further problem that no patient-level adjustment solves. Length of stay, discharge destination and readmission depend on rehabilitation capacity, reimbursement rules, the practice of transferring patients back to referring hospitals and the availability of community care, all of which differ between countries and regions. A common international TO definition can standardize the components and their windows, but TO rates should be compared within a health system, and cross-national comparison should be limited to the death and major morbidity components until the effects of the system on stay and readmission have been quantified. Within a system, three patient characteristics deserve priority in the adjustment model because they act on the components that fail most often: frailty, preoperative functional status and preoperative living situation [71]. Discharge destination is an outcome and not a risk factor, and it belongs among the supplementary indicators, with the caution that the STS response categories changed between data versions and that a discharge to a facility for

rehabilitation with the intention of returning home is coded differently from a discharge to a nursing home [27].

9. Evolving Paradigms

Four developments will shape how TO is used. The first is registry embedding, which is less complete than a component list suggests. The STS Adult Cardiac Surgery Database contains a field for every Tier 1 component proposed in Section 10, including readmission to any acute care hospital within 30 days of discharge [27]. The EACTS Adult Cardiac Database data specification contains mortality, stroke, new dialysis, sternal wound infection, reoperation and a stay above 14 days, but defines prolonged ventilation as more than 48 h and has no readmission field [28]. The NHR registers mortality and major complications, and its published description presents length of stay and readmission as information to be gained by linkage with insurance and hospital discharge data [104]. Readmission is therefore the component that most often requires linkage, and both routes to it are imperfect. Site abstraction misses readmissions to other hospitals. When 45,539 STS records were matched to regional billing data, the registry-reported readmission rate was 10.5% and the true rate 13.0%, and about 36% of readmissions were to a different hospital system [105]. In the NRD, 23% of 30-day readmissions after CABG and 26% after surgical aortic valve replacement were to a hospital other than the index hospital, and in New South Wales only 35.7% of readmitted patients who had not been transferred before surgery returned to the operating hospital [106,107]. Administrative linkage captures these events but has its own limits. Linkage is incomplete and selective: 86% of eligible Medicare records could be linked to STS records, and the linked measure covers Medicare beneficiaries only [57]. Claims databases that follow patients within one state or one insurer miss readmissions outside it. Diagnosis coding undercounts complications and varies between hospitals [43]. Linkage is governed by national privacy law and is not available in every country, and where it is unavailable, the readmission component cannot be scored. A TO report should therefore state how readmission was ascertained, the proportion of records linked and the proportion of patients with any missing component, and should present a TO rate without the readmission component whenever capture is incomplete.

The second development is the shift of enhanced recovery from a set of protocols to a measured outcome. The Enhanced Recovery After Surgery Society has issued recommendations for perioperative care in cardiac surgery, and TO or DAH30 could serve as the outcome against which a program is evaluated, although no study has yet used TO in this way [78]. The third is prediction. In a meta-analysis of 15 studies, machine learning models discriminated operative mortality after cardiac surgery better than logistic regression, with pooled c-statistics of 0.88 and 0.81 [108]. Whether the same holds for TO is unknown. A model that identified before surgery the patients likely to fail on stay or readmission could direct prehabilitation and discharge planning, but it would need the calibration and external validation described in Section 6 before use. The fourth is the move from binary to graded outcomes through DOOR, the win ratio or DAH30 [75,76,79]. Graded measures may replace TO for statistical comparison, and TO may remain useful as the summary that is communicated to patients and administrators. Which of these happens depends on the head-to-head evidence and on whether the specialty can agree on an outcome hierarchy.

10. A Proposed Preliminary Core Set for Adult Cardiac Surgery and a Research Agenda

The controversies above lead to a proposal, which we present as a preliminary framework for discussion (Figure 3, Table 4, Table S2). It is the authors' proposal. It has not been

through a consensus process, has had no patient or public input, and has not been tested in any dataset; no tier, threshold or reporting rule in it has been validated. The framework has one composite and two sets of supplementary indicators. Tier 1 is the proposed core: a single binary, all-or-none measure built from fields of the STS Adult Cardiac Surgery Database in their current definitions. A patient achieves Tier 1 TO with no operative mortality, no permanent stroke, no renal failure, no prolonged ventilation above 24 h, no reoperation for a cardiac reason, no deep sternal wound infection or mediastinitis, a postoperative stay of 14 days or less, and no readmission to an acute care hospital within 30 days of discharge. The complication list is the STS major morbidity set, and four of its five items are the complications of the STS FTR metric, so that TO and FTR are computed from the same events [1,8,9]. The length-of-stay criterion is the existing STS and EACTS definition of a prolonged stay, an absolute threshold that keeps its meaning over time and uses the same 14-day cutoff as the national TO cohort [9,16,28]. Published STS event rates allow only an upper bound on the result. In the STS model development sample, the composite of death or major morbidity occurred in 15.0% of CABG, 18.4% of valve and 28.3% of combined valve and CABG operations, so Tier 1 TO could not have exceeded 85.0%, 81.6% and 71.7% in those groups in that period, before stay and readmission were counted [9]. In 2024, operative mortality after isolated CABG was 2.1% and the 30-day readmission rate was 9.5% [109]. A registry-based Tier 1 rate will therefore be lower than the 86% of the administrative cohort and will differ by procedure group, and the two should not be compared.

Tiers 2 and 3 are not composites. Each of their elements is a supplementary indicator that is reported individually, with its own denominator and its own completeness, and none is combined with another or folded into Tier 1. A hospital therefore has one Tier 1 rate and a short list of separately reported indicators. Tier 2 holds four patient-centered indicators: the proportion of patients discharged home; DAH30 as a continuous measure, reported as the median with interquartile range and the proportion of patients with 0 days; the proportion with a new permanent pacemaker or defibrillator after valve procedures; and a patient-reported health status measure, reported as the change from baseline at 3 months with its response rate. Tier 3 holds 1-year durability indicators: all-cause survival, freedom from cardiac reintervention and, for valve procedures, freedom from valve-related events as defined in the societies' reporting guidelines and by the Valve Academic Research Consortium (structural valve deterioration, valve thrombosis, embolism, major bleeding and prosthetic valve endocarditis) [110,111]. Each Tier 3 indicator is estimated by the Kaplan-Meier method from the date of surgery, the source of vital status and event data (national death index, population register or active follow-up) is stated, and the proportion of patients with complete 1-year follow-up is reported beside it, because an indicator ascertained in a minority of patients describes follow-up and not durability. Tier 1 is intended for adult CABG, valve and combined operations. It is reported separately for isolated CABG, isolated valve and combined valve and CABG procedures, and separately for elective and urgent operations. Emergency and salvage operations are reported descriptively and are excluded from hospital comparison until a model calibrated in those patients exists, since the registry evidence summarized here comes from elective cohorts. Components that are specific to one procedure, such as a new pacemaker after valve surgery, stay outside Tier 1 so that the core definition is the same for every procedure.

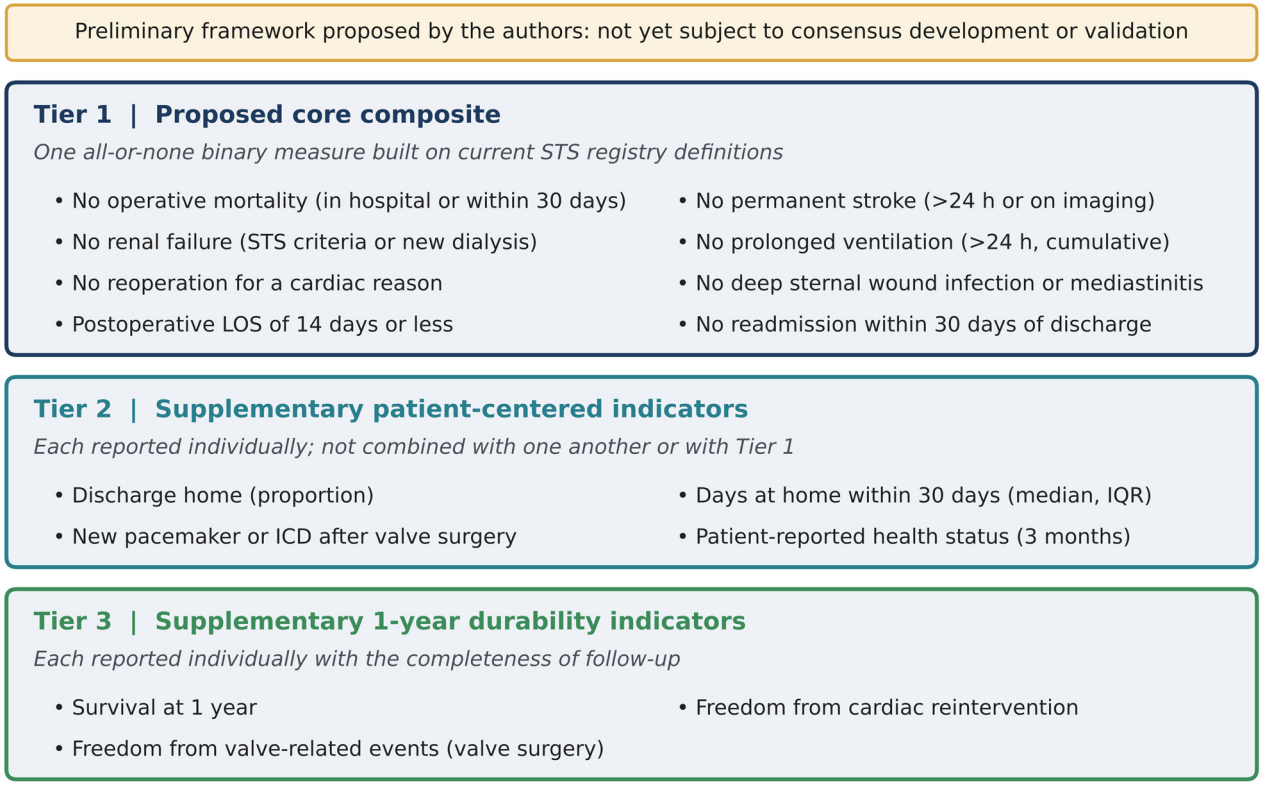


Figure 3. Proposed preliminary framework for textbook outcome reporting in adult cardiac surgery. Tier 1 is a single all-or-none composite built from current Society of Thoracic Surgeons registry definitions. Tier 2 (patient-centered) and Tier 3 (1-year durability) contain supplementary indicators that are each reported individually and are not combined with one another or with Tier 1. The reporting rules at the foot of the figure apply to every textbook outcome report. The framework is the authors' proposal and has not undergone consensus development or validation. FTR, failure to rescue; ICD, implantable cardioverter defibrillator; IQR, interquartile range; LOS, length of stay; STS, Society of Thoracic Surgeons; TO, textbook outcome.

Two scoring rules of Tier 1 need to be stated exactly, because each changes which patients count as failures. The first concerns readmission. The STS field records any inpatient readmission to an acute care hospital within 30 days of discharge, whether planned or unplanned and whether or not related to the operation, and Tier 1 follows it [27]. A planned readmission, for example, for a non-cardiac operation that was deferred until the cardiac condition had been treated or for a staged procedure, is scheduled care and not an unfavorable outcome. Publicly reported claims-based readmission measures exclude such admissions for that reason; the algorithm developed for the purpose classified 7.8% of readmissions in a hospital-wide cohort as planned, and its revised version had a positive predictive value of 58.7% against chart review [112]. The national TO cohort likewise analyzed non-elective readmissions only [16]. Counting planned readmissions therefore scores some patients with an uneventful course as failures, and it lowers the TO rate of hospitals that stage procedures or that operate on patients who are being prepared for

another intervention. The size of this effect in cardiac surgery has not been reported in the studies reviewed here. We keep the registry definition in the core, because it is the field that registries hold and because it requires no judgment about intent. The interpretation is protected in another way. The STS reason field has a category for planned readmissions, defined as readmission for a procedure that was conditional on surgical treatment of the cardiac condition [27]. Readmissions recorded in that category are counted separately in the component profile, and Tier 1 is also reported with them excluded, so that a reader can see how much of a hospital's readmission failure is scheduled care.

The second rule concerns missing data, and the choice affects hospital comparison. In the primary analysis, every eligible patient remains in the denominator. A patient with at least one observed failed component has not achieved TO, whatever else is missing. A patient with no observed failure and at least one missing component has an undetermined status and is also classified as not achieving TO; such patients are not excluded from the denominator. Excluding them is a complete-case analysis, which can be biased unless the data are missing completely at random, and it would raise the TO rate of a hospital whose records are more often incomplete after a complicated course [113]. The primary rule errs in the opposite direction: it lowers the TO rate of a hospital with incomplete records whatever its outcomes, so that a low rate may reflect data capture and not care. The two approaches can diverge widely, since 30-day mortality status was missing or unknown in 15.0% of records in the STS model development sample [9]. Three safeguards follow. The proportion of patients with undetermined status is reported for every hospital alongside its TO rate. A hospital in which that proportion exceeds a prespecified limit is reported but is not compared with other hospitals. The analysis is repeated with undetermined patients excluded from the denominator, with multiple imputation, and with the incompletely captured component removed, and a hospital whose classification differs between these analyses is not classified (Table S2).

Five reporting rules accompany the set and are equally provisional. Every TO rate is reported with its component profile: the proportion of patients failing each component, the proportion failing each component alone and the distribution of the number of failed components. Risk adjustment uses a model developed and validated for TO as described in Section 6, and until one exists, the model used and its calibration for TO are stated. Hospital comparison uses a hierarchical model, reports the risk-standardized rate with a 95% interval, the between-hospital variance and the reliability, and classifies no hospital whose reliability is below the prespecified minimum. Clinical-risk-adjusted TO is reported stratified by social risk, with a social-risk-adjusted rate for accountability uses. TO is reported beside risk-adjusted FTR on the same cohort and period, and beside mortality and the STS composite where these are available, so that it supplements and does not replace them. The component profile also shows the number of readmissions recorded as planned and the number of patients with undetermined status, as set out above.

The research agenda follows from the gaps identified in each section, and Table 5 separates what the published evidence establishes from what we infer and from what remains untested. The first priority is a head-to-head comparison of Tier 1 TO, the STS composite and risk-adjusted FTR in one registry, reporting agreement, reclassification, reliability and prediction of subsequent outcomes. The second is a TO-specific risk model that includes frailty, functional status and social risk, with external validation. The third is an empirical study of reliability and minimum caseload, with sensitivity analyses that remove length of stay and that replace the binary measure with ordinal and continuous outcomes. The fourth is a Delphi consensus on the core definition under the auspices of the STS and EACTS, with patients and caregivers as a stakeholder group, following the example set in gastrectomy [55]. The fifth is a prospective evaluation in a regional collaborative of whether feedback of TO with FTR changes processes or outcomes, with the

Michigan collaborative as a model for the infrastructure [38,77]. The sixth is validation of the patient-reported and days-at-home indicators against each other and against 1-year outcomes. Most of these studies can be done with existing data. Until they are done, TO in adult cardiac surgery is a candidate measure.

Table 4. Proposed preliminary textbook outcome framework for adult cardiac surgery: elements, operational definitions, ascertainment windows, registry sources and rationale.

Element	Operational Definition	Ascertainment Window	Registry Source	Rationale
Tier 1: proposed core composite (one all-or-none binary measure)				
Operative mortality	Death from any cause during the index hospitalization, including after transfer to another acute care facility and including discharge to hospice, or death after discharge within 30 days of the operation	Index hospitalization of any duration, or 30 days from surgery	STS operative mortality; EACTS 30-day mortality and status at discharge	Single STS definition; the same event as in the STS composite and FTR
Permanent stroke	New neurological deficit of abrupt onset caused by a disturbance in cerebral blood supply that is confirmed on imaging or does not resolve within 24 h	Index hospitalization	STS; EACTS (24 h or more, or imaging)	STS major morbidity and FTR complication
Renal failure	Creatinine at least 3 times the preoperative value, or 4.0 mg/dL or more with a rise of at least 0.5 mg/dL, or new dialysis	Index hospitalization	STS; EACTS records new dialysis only	STS major morbidity and FTR complication
Prolonged ventilation	More than 24 h of postoperative ventilation, counted from operating room exit to extubation plus any hours after reintubation	Index hospitalization	STS; EACTS threshold is more than 48 h	STS major morbidity and FTR complication
Reoperation	Re-exploration for bleeding or tamponade, reintervention for valve dysfunction, unplanned coronary or aortic intervention, or unplanned return to the operating room for another cardiac reason; planned staged procedures, device removal, pacemaker implantation and non-cardiac procedures are not counted	Index hospitalization	STS; EACTS reoperation fields	STS major morbidity and FTR complication (definition of the 2018 models)
Deep sternal wound infection or mediastinitis	Deep incisional or organ-space infection meeting the STS criteria	30 days from surgery or index hospitalization	STS; EACTS	STS major morbidity; not an FTR complication
Postoperative length of stay	Discharge from acute care within 14 days of the operation; transferred patients are followed to final acute care discharge	Surgery to acute care discharge	STS prolonged stay (more than 14 days); EACTS long stay field	Absolute threshold already in both registries; percentile thresholds reserved for sensitivity analysis
Readmission	No inpatient admission to any acute care hospital, planned or unplanned, within 30 days of discharge; observation stays and admissions to rehabilitation or skilled nursing are not counted; readmissions recorded as planned are shown separately in the component profile and are excluded in a sensitivity analysis	30 days from discharge	STS readmission field, preferably verified by administrative linkage; not collected by EACTS	Patient-relevant; not part of the STS composite [27,58]
Tier 2: supplementary patient-centered indicators (each reported individually; not combined)				
Discharge home	Discharge to the home of the patient, including assisted living or the home of a relative	At discharge	STS discharge location	Proxy for functional recovery used in aortic TO [23]
DAH30	30 minus postoperative days in the index hospitalization and days of acute readmission within 30 days of surgery; institution	30 days from surgery	Registry dates plus linkage	Continuous; sensitivity outcome for TO [76]

	days subtracted when recorded; 0 if the patient dies within 30 days			
New permanent pacemaker or defibrillator (valve procedures)	New dysrhythmia requiring insertion of a permanent device	Index hospitalization	STS; EACTS	Procedure-specific and patient-relevant
Patient-reported health status	7-item Seattle Angina Questionnaire for coronary surgery; EQ-5D-5L with a valve-specific or heart failure instrument for valve surgery; change from baseline with response rate	Baseline and 3 months	Prospective collection	Patient-assessed outcome; instruments from the ICHOM sets [92,94]
Tier 3: supplementary 1-year durability indicators (each reported individually with follow-up completeness)				
Survival at 1 year	Death from any cause; Kaplan-Meier estimate	365 days from surgery	Linkage to a death index or population register	Durability; a component of mitral and transplant TO [17,45]
Freedom from cardiac reintervention	Any repeat surgical or transcatheter procedure on the operated valve or on the revascularized coronary territory	365 days from surgery	Registry follow-up or linkage	Durability of the index operation
Freedom from valve-related events (valve procedures)	Structural valve deterioration, valve thrombosis, embolism, major bleeding and prosthetic valve endocarditis	365 days from surgery	Active follow-up with echocardiography	Definitions from the reporting guidelines [110,111]

All elements are proposed by the authors and have not undergone consensus development or validation. Tier 1 is a single all-or-none composite reported with its component profile. Tier 2 and Tier 3 elements are individual indicators and are not combined with one another or with Tier 1. Definitions follow the STS Adult Cardiac Surgery Database, data version 4.20.2 [27], and are compared with the EACTS Adult Cardiac Database data specification, version 3.3 [28]; exclusions and the handling of missing data are given in Table S2. In the primary analysis, a patient with a missing component and no observed failure remains in the denominator and is classified as not achieving TO (Section 10). DAH30, days at home within 30 days of surgery; EACTS, European Association for Cardio-Thoracic Surgery; FTR, failure to rescue; ICHOM, International Consortium for Health Outcomes Measurement; STS, Society of Thoracic Surgeons; TO, textbook outcome.

Table 5. Established evidence, authors' interpretation and research needed for each question addressed in this review.

Question	What Published Studies Establish	Authors' Interpretation or Proposal	Evidence Needed
Can TO be measured in adult cardiac surgery?	TO has been calculated from administrative and registry data: 86.0% and 87.1% in two elective cohorts, 59.1% to 81.9% in two coronary series [16–19]	The rates are not comparable because the definitions differ	One consensus definition applied in more than one registry
Is TO associated with later outcomes?	In one adult cardiac cohort TO was associated with lower 180-day mortality and readmission [16]; associations with survival were reported in transplant, congenital and endovascular aortic cohorts [22,24,45–48]	A patient-level association does not show that hospital TO rates measure quality	Hospital TO rates predicting subsequent hospital outcomes
Does TO vary between hospitals?	The ICC for TO was 8.5% in one cohort, against 14.5% for in-hospital mortality and FTR [16]; reliability was not reported	A high prevalence need not preclude reliable measurement (illustration in Note S1)	Reliability and minimum caseload estimated from registry data
Is length of stay the main determinant?	Prolonged stay was present in 58.6% of failures in one administrative cohort and was the only recorded failure in about one third [16]	In a registry-based definition major morbidity is likely to be more frequent than a stay above 14 days	Component profile of Tier 1 TO in registry data, by procedure group

Is risk adjustment adequate?	One cohort adjusted with a multilevel administrative model without reporting its performance [16]; no calibrated or validated TO model has been published	Mortality models are likely to under-adjust for stay and readmission	Development, calibration and external validation of a TO model
Does TO add to the STS composite or to FTR?	No comparison with the STS composite. Hospital TO rates correlated weakly with in-hospital mortality (r = −0.37) and with FTR (r = −0.09) in one administrative cohort [16]; TO and FTR were evaluated as a pair in one aortic registry study [25]	TO may identify hospitals with slow recovery or readmission without major morbidity; pairing with FTR may localize problems	Head-to-head comparison in one registry; evaluation of feedback
Is TO patient-centered?	Patients preferred a summary TO format in one non-cardiac survey [83]; no cardiac definition reports patient involvement	The components are plausible patient priorities but are unconfirmed	Consensus process with patient participation; validation of patient-reported measures
Does TO reflect social risk?	TO is associated with income, insurance and neighborhood or county-level social measures in administrative data [16,20,21,84]	Stratified and adjusted results should be reported together	Effect of social risk adjustment on hospital TO classification
Does measuring TO improve care?	No evidence	No claim is made	Prospective evaluation in a quality collaborative

Statements in the second column are limited to what the cited studies report. Statements in the third column are the authors' inferences or proposals and are not established. FTR, failure to rescue; ICC, intraclass correlation coefficient; STS, Society of Thoracic Surgeons; TO, textbook outcome.

11. Limitations

This review has limitations that bear on every conclusion. It is narrative. The search was structured and is reproducible, but it used one database, PubMed, and other databases, conference proceedings and the gray literature were not searched, so TO studies indexed elsewhere may have been missed. No language limit was applied, yet all retrieved records were in English, and studies published in other languages or in journals not indexed in MEDLINE are not represented. A preliminary search restricted to the Title field missed four adult cardiac studies that the final search found, which shows that TO is often reported as a secondary outcome and that further studies may exist in which it is not named in the title or abstract. Screening was not performed independently in duplicate. We did not assess risk of bias in the included studies, all of which are observational and most of which are retrospective analyses of administrative or registry data, and we did not pool TO rates or test for publication bias; studies in which TO was unrelated to outcome or did not vary between hospitals may be less likely to have been published. Component-level data were available for only a minority of studies, and several entries in Table 1 rest on the published summary of the study.

The evidence itself is limited. The adult cardiac literature consists of two registry analyses from two countries, two coronary series from one group and two claims-based analyses, all with different definitions, and the registry evidence concerns elective surgery. The comparator studies in transplantation, congenital and aortic surgery inform the methodological discussion only. The reliability figures in Section 6 are an illustration under stated assumptions and not an analysis of data. The proposed framework reflects the authors' judgment and the structure of the STS registry. It has had no formal consensus process, no patient or public involvement and no external validation; its 14-day threshold and its ascertainment windows are inherited from existing fields and not derived from evidence of what patients or clinicians consider ideal; and its applicability to registries with different fields, to urgent and emergency surgery and to health systems with

different discharge pathways is unknown. Finally, members of our group have published a systematic review of TO in cardiothoracic surgery and two reviews of FTR, three of the references in this article are our own, and TO is the subject of the first author's doctoral research [15,41,42]. The pairing of TO with FTR is our framework, and readers should weigh our interpretation accordingly. We have tried to mark it as interpretation wherever it appears.

12. Conclusions

TO has entered the adult cardiac surgical literature through two registry analyses and a small number of single-group and claims-based studies. They show that TO can be calculated from existing data, that it was achieved in 86% to 87% of elective patients under two different definitions and in 59% to 82% under a stricter one, that prolonged hospitalization was the most frequent reason for failure in the one cohort that reported it, and that TO was associated with lower 180-day mortality and readmission in that cohort, where hospital TO rates also correlated only weakly with in-hospital mortality and FTR. They do not show that TO measures hospital quality better than, or in addition to, the STS composite score, that hospital TO rates are reliable at ordinary caseloads, that existing risk models are adequate for it, or that measuring it improves care. TO is at present a promising and easily understood candidate measure without a standard definition, a validated risk model or comparative evidence. We have proposed a preliminary registry-based core definition, supplementary patient-centered and durability indicators, paired reporting with FTR and an analytic framework, and we have set out the studies that would confirm or refute its usefulness. The next steps are a formal consensus on the definition, with patient participation, and a head-to-head evaluation in registry data. Until those are complete, TO is suited to research and local quality improvement and not to public reporting or payment.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jcdd13120618/s1, Table S1: Search strategy and classification of the 43 screened records; Table S2: Operational definitions, ascertainment windows, data sources, exclusions and handling of missing data for each element of the proposed preliminary framework; Note S1: Illustrative reliability calculation for hospital-level textbook outcome rates.

Author Contributions: Conceptualization, D.E.M., B.R. and A.X.; methodology, D.E.M. and S.S.; literature search and data curation, D.E.M., V.A. and P.-A.Z.; writing, original draft preparation, D.E.M.; writing, review and editing, S.S., V.A., P.-A.Z., B.R., D.Z. and A.X.; visualization, D.E.M.; supervision, B.R., D.Z. and A.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new patient-level data were collected or generated for this review, and no dataset was created. The confidence intervals, derived proportions and illustrative reliability estimates presented in the article are secondary calculations made by the authors from aggregate data in the cited publications. The source of each input and the method of calculation are given in Section 2, in the footnote of Table 1, in the legend of Figure 2 and in Supplementary Note S1, and the records identified by the search are listed in Table S1. No other data are available, and data sharing is otherwise not applicable to this article.

Acknowledgments: Figures 1, 2, 3, and the Graphical Abstract were created in BioRender. Magouliotis, D. (2026) https://BioRender.com/fsq88wc; https://BioRender.com/bf18bl7; https://BioRender.com/ina733a; https://BioRender.com/mbzx4ii.

Conflicts of Interest: The authors declare no financial conflicts of interest. The authors declare the following non-financial, intellectual interests: five of the authors (D.E.M., V.A., P.-A.Z., D.Z. and A.X.) are co-authors of a previously published systematic review of textbook outcome in cardiothoracic surgery [15], four of the authors (D.E.M., V.A., P.-A.Z. and A.X.) are co-authors of previously published reviews of failure to rescue [41,42], and textbook outcome is the subject of the doctoral research of D.E.M. The framework proposed in this article builds on that work, and the authors therefore have a scholarly interest in its adoption.

Abbreviations

The following abbreviations are used in this manuscript:

AAA	Abdominal aortic aneurysm
CABG	Coronary artery bypass grafting
CI	Confidence interval
DAH30	Days at home within 30 days of surgery
DAOH	Days alive and out of hospital
DOOR	Desirability of outcome ranking
DSAA	Dutch Surgical Aneurysm Audit
EACTS	European Association for Cardio-Thoracic Surgery
EVAR	Endovascular aortic repair
FTR	Failure to rescue
ICC	Intraclass correlation coefficient
ICHOM	International Consortium for Health Outcomes Measurement
LOS	Length of stay
NHR	Netherlands Heart Registration
NRD	Nationwide Readmissions Database
STS	Society of Thoracic Surgeons
TEVAR	Thoracic endovascular aortic repair
TO	Textbook outcome
UNOS	United Network for Organ Sharing
VQI	Vascular Quality Initiative

References

1. Kurlansky, P.A.; O'Brien, S.M.; Vassileva, C.M.; Lobdell, K.W.; Edwards, F.H.; Jacobs, J.P.; Wyler von Ballmoos, M.; Paone, G.; Edgerton, J.R.; Thourani, V.H.; et al. Failure to Rescue: A New Society of Thoracic Surgeons Quality Metric for Cardiac Surgery. *Ann. Thorac. Surg.* **2022**, *113*, 1935–1942. https://doi.org/10.1016/j.athoracsur.2021.06.025.
2. Nashef, S.A.; Roques, F.; Sharples, L.D.; Nilsson, J.; Smith, C.; Goldstone, A.R.; Lockowandt, U. EuroSCORE II. *Eur. J. Cardiothorac. Surg.* **2012**, *41*, 734–744. https://doi.org/10.1093/ejcts/ezs043.
3. Dimick, J.B.; Staiger, D.O.; Birkmeyer, J.D. Ranking hospitals on surgical mortality: The importance of reliability adjustment. *Health Serv. Res.* **2010**, *45*, 1614–1629. https://doi.org/10.1111/j.1475-6773.2010.01158.x.
4. Shahian, D.M.; Edwards, F.H.; Ferraris, V.A.; Haan, C.K.; Rich, J.B.; Normand, S.L.; DeLong, E.R.; O'Brien, S.M.; Shewan, C.M.; Dokholyan, R.S.; et al. Quality measurement in adult cardiac surgery: Part 1: Conceptual framework and measure selection. *Ann. Thorac. Surg.* **2007**, *83*, S3–S12. https://doi.org/10.1016/j.athoracsur.2007.01.053.
5. O'Brien, S.M.; Shahian, D.M.; DeLong, E.R.; Normand, S.L.; Edwards, F.H.; Ferraris, V.A.; Haan, C.K.; Rich, J.B.; Shewan, C.M.; Dokholyan, R.S.; et al. Quality measurement in adult cardiac surgery: Part 2: Statistical considerations in composite measure scoring and provider rating. *Ann. Thorac. Surg.* **2007**, *83*, S13–S26. https://doi.org/10.1016/j.athoracsur.2007.01.055.
6. Shahian, D.M.; O'Brien, S.M.; Normand, S.L.; Peterson, E.D.; Edwards, F.H. Association of hospital coronary artery bypass volume with processes of care, mortality, morbidity, and the Society of Thoracic Surgeons composite quality score. *J. Thorac. Cardiovasc. Surg.* **2010**, *139*, 273–282. https://doi.org/10.1016/j.jtcvs.2009.09.007.

7. Shahian, D.M.; He, X.; Jacobs, J.P.; Rankin, J.S.; Welke, K.F.; Filardo, G.; Shewan, C.M.; O'Brien, S.M. The Society of Thoracic Surgeons Isolated Aortic Valve Replacement (AVR) Composite Score: A report of the STS Quality Measurement Task Force. *Ann. Thorac. Surg.* **2012**, *94*, 2166–2171. https://doi.org/10.1016/j.athoracsur.2012.08.120.
8. Shahian, D.M.; Jacobs, J.P.; Badhwar, V.; Kurlansky, P.A.; Furnary, A.P.; Cleveland, J.C.; Lobdell, K.W.; Vassileva, C.; Wyler von Ballmoos, M.C.; Thourani, V.H.; et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1-Background, Design Considerations, and Model Development. *Ann. Thorac. Surg.* **2018**, *105*, 1411–1418. https://doi.org/10.1016/j.athoracsur.2018.03.002.
9. O'Brien, S.M.; Feng, L.; He, X.; Xian, Y.; Jacobs, J.P.; Badhwar, V.; Kurlansky, P.A.; Furnary, A.P.; Cleveland, J.C.; Lobdell, K.W.; et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results. *Ann. Thorac. Surg.* **2018**, *105*, 1419–1428. https://doi.org/10.1016/j.athoracsur.2018.03.003.
10. Kolfschoten, N.E.; Kievit, J.; Gooiker, G.A.; van Leersum, N.J.; Snijders, H.S.; Eddes, E.H.; Tollenaar, R.A.; Wouters, M.W.; Marang-van de Mheen, P.J. Focusing on desired outcomes of care after colon cancer resections; hospital variations in 'textbook outcome'. *Eur. J. Surg. Oncol.* **2013**, *39*, 156–163. https://doi.org/10.1016/j.ejso.2012.10.007.
11. Karthaus, E.G.; Lijftogt, N.; Busweiler, L.A.D.; Elsman, B.H.P.; Wouters, M.W.J.M.; Vahl, A.C.; Hamming, J.F. Textbook Outcome: A Composite Measure for Quality of Elective Aneurysm Surgery. *Ann. Surg.* **2017**, *266*, 898–904. https://doi.org/10.1097/SLA.0000000000002388.
12. Busweiler, L.A.; Schouwenburg, M.G.; van Berge Henegouwen, M.I.; Kolfschoten, N.E.; de Jong, P.C.; Rozema, T.; Wijnhoven, B.P.; van Hillegersberg, R.; Wouters, M.W.; van Sandick, J.W.; et al. Textbook outcome as a composite measure in oesophagogastric cancer surgery. *Br. J. Surg.* **2017**, *104*, 742–750. https://doi.org/10.1002/bjs.10486.
13. van Roessel, S.; Mackay, T.M.; van Dieren, S.; van der Schelling, G.P.; Nieuwenhuijs, V.B.; Bosscha, K.; van der Harst, E.; van Dam, R.M.; Liem, M.S.L.; Festen, S.; et al. Textbook Outcome: Nationwide Analysis of a Novel Quality Measure in Pancreatic Surgery. *Ann. Surg.* **2020**, *271*, 155–162. https://doi.org/10.1097/SLA.0000000000003451.
14. Ten Berge, M.G.; Beck, N.; Steup, W.H.; Verhagen, A.F.T.M.; van Brakel, T.J.; Schreurs, W.H.; Wouters, M.W.J.M. Textbook outcome as a composite outcome measure in non-small-cell lung cancer surgery. *Eur. J. Cardiothorac. Surg.* **2021**, *59*, 92–99. https://doi.org/10.1093/ejcts/ezaa265.
15. Magouliotis, D.E.; Androutsopoulou, V.; Zotos, P.A.; Cioffi, U.; Minervini, F.; Sicouri, N.; Zacharoulis, D.; Xanthopoulos, A.; Scarci, M. Beyond Mortality: Textbook Outcome as a Novel Quality Metric in Cardiothoracic Surgical Care. *J. Clin. Med.* **2025**, *14*, 7660. https://doi.org/10.3390/jcm14217660.
16. Verma, A.; Balian, J.; Sakowitz, S.; Sanaiha, Y.; Benharash, P. Textbook outcome as a hospital benchmarking metric in adult cardiac surgery. *Surgery* **2026**, *196*, 110311. https://doi.org/10.1016/j.surg.2026.110311.
17. Ko, K.; Heuts, S.; Tjon Joek Tjien, A.; Sardari Nia, P.; Maessen, J.G.; Houterman, S.; Singh, S.K.; Boon, R.; Tonino, P.; Verberkmoes, N.; et al. Textbook Outcome in Mitral Valve Surgery-Results from the Netherlands Heart Registration. *Eur. J. Cardiothorac. Surg.* **2025**, *67*, ezaf422. https://doi.org/10.1093/ejcts/ezaf422.
18. Görtzen, Q.; Sampon, F.; Timmermans, N.; Ter Woorst, J.; Akca, F. Endoscopic-assisted, minimally invasive versus sternotomy total arterial multivessel bypass grafting. *Interdiscip. Cardiovasc. Thorac. Surg.* **2024**, *39*, ivae187. https://doi.org/10.1093/icvts/ivae187.
19. Sampon, F.; Ter Woorst, J.; Dekker, L.; Akca, F. Thoracoscopic-assisted, minimally invasive versus off-pump bypass grafting for single vessel coronary artery disease—A propensity matched analysis. *Int. J. Cardiol.* **2024**, *409*, 132175. https://doi.org/10.1016/j.ijcard.2024.132175.
20. Azap, L.; Woldesenbet, S.; Osae, H.; Henn, M.; Bozinovski, J.; Whitson, B.; Mokadam, N.A.; Ganapathi, A.; Pawlik, T.M. County-level food insecurity is associated with outcomes following cardiac surgery. *Am. J. Surg.* **2024**, *237*, 115907. https://doi.org/10.1016/j.amjsurg.2024.115907.
21. Chen, Y.; Xiao, Y.; Huang, R.; Jiang, F.; Zhou, J.; Su, C.; Yang, T. Association between hospital racial composition and aortic valve replacement outcomes: A national inpatients sample database analysis. *Catheter. Cardiovasc. Interv.* **2024**, *103*, 637–649. https://doi.org/10.1002/ccd.30970.
22. Cook, J.K.; Ghandour, H.; Singh, E.S.; Foster, J.A.; Prabhu, N.K.; Mensah-Mamfo, M.; Moya-Mendez, M.E.; Medina, C.K.; Thornton, S.W.; Overbey, D.M.; et al. Textbook Outcome for the Fontan Operation: A Holistic Quality Metric in Congenital Heart Surgery. *Innovations* **2025**, *20*, 464–471. https://doi.org/10.1177/15569845251375441.
23. Felsted, A.; Beck, A.W.; Banks, C.A.; Neal, D.; Columbo, J.A.; Robinson, S.T.; Stone, D.H.; Scali, S.T. A patient-centered textbook outcome measure effectively discriminates contemporary elective open abdominal aortic aneurysm repair quality. *J. Vasc. Surg.* **2024**, *80*, 1071–1081.e5. https://doi.org/10.1016/j.jvs.2024.05.052.

24. Sickels, A.D.; Banks, C.A.; Filiberto, A.C.; Novak, Z.; Spangler, E.L.; Rowse, J.; Motyl, C.M.; Scali, S.T.; Scully, R.E.; Columbo, J.A.; et al. "Textbook Outcome" in Endovascular Aortic Repair is Associated With Improved One-year Survival. *Ann. Surg.* **2026**, *283*, 878–887. https://doi.org/10.1097/SLA.0000000000006590.
25. Lentz, C.M.; Mani, K.; von Meijenfeldt, G.C.I.; van der Laan, M.J.; Wanhainen, A.; Asciutto, G.; Zeebregts, C.J. Feasibility and Added Value of Textbook Outcome and Failure to Rescue in Elective Juxtarenal Abdominal Aortic Aneurysm Repair: Insights from Two National Registries. *Eur. J. Vasc. Endovasc. Surg.* **2026**, *in press*. https://doi.org/10.1016/j.ejvs.2026.01.039.
26. Baethge, C.; Goldbeck-Wood, S.; Mertens, S. SANRA-a scale for the quality assessment of narrative review articles. *Res. Integr. Peer Rev.* **2019**, *4*, 5. https://doi.org/10.1186/s41073-019-0064-8.
27. The Society of Thoracic Surgeons. *STS Adult Cardiac Surgery Database Training Manual: Data Specifications Version 4.20.2 (Updated May 2021)*; The Society of Thoracic Surgeons: Chicago, IL, USA, 2021. Available online: https://www.sts.org/sites/default/files/ACSD%20Training%20Manual%20V4_20_2%20May2021.pdf (accessed on 9 November 2026).
28. European Association for Cardio-Thoracic Surgery. *EACTS Adult Cardiac Database, Quality Improvement Programme (QUIP): Data Specification*, Version 3.3; EACTS: Windsor, UK, 2026. Available online: https://www.eacts.org/wp-content/uploads/2026/09/EACTS_DataSpec_ACD_v3.3.xlsx (accessed on 10 November 2026).
29. Donabedian, A. The quality of care. How can it be assessed? *JAMA* **1988**, *260*, 1743–1748. https://doi.org/10.1001/jama.260.12.1743.
30. Nolan, T.; Berwick, D.M. All-or-none measurement raises the bar on performance. *JAMA* **2006**, *295*, 1168–1170. https://doi.org/10.1001/jama.295.10.1168.
31. Staiger, D.O.; Dimick, J.B.; Baser, O.; Fan, Z.; Birkmeyer, J.D. Empirically derived composite measures of surgical performance. *Med. Care* **2009**, *47*, 226–233. https://doi.org/10.1097/MLR.0b013e3181847574.
32. Dimick, J.B.; Staiger, D.O.; Hall, B.L.; Ko, C.Y.; Birkmeyer, J.D. Composite measures for profiling hospitals on surgical morbidity. *Ann. Surg.* **2013**, *257*, 67–72. https://doi.org/10.1097/SLA.0b013e31827b6be6.
33. Reeves, D.; Campbell, S.M.; Adams, J.; Shekelle, P.G.; Kontopantelis, E.; Roland, M.O. Combining multiple indicators of clinical quality: An evaluation of different analytic approaches. *Med. Care* **2007**, *45*, 489–496. https://doi.org/10.1097/MLR.0b013e31803bb479.
34. Shwartz, M.; Restuccia, J.D.; Rosen, A.K. Composite Measures of Health Care Provider Performance: A Description of Approaches. *Milbank Q.* **2015**, *93*, 788–825. https://doi.org/10.1111/1468-0009.12165.
35. Ferreira-González, I.; Busse, J.W.; Heels-Ansdell, D.; Montori, V.M.; Akl, E.A.; Bryant, D.M.; Alonso-Coello, P.; Alonso, J.; Worster, A.; Upadhye, S.; et al. Problems with use of composite end points in cardiovascular trials: Systematic review of randomised controlled trials. *BMJ* **2007**, *334*, 786. https://doi.org/10.1136/bmj.39136.682083.AE.
36. Silber, J.H.; Williams, S.V.; Krakauer, H.; Schwartz, J.S. Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. *Med. Care* **1992**, *30*, 615–629. https://doi.org/10.1097/00005650-199207000-00004.
37. Ghaferi, A.A.; Birkmeyer, J.D.; Dimick, J.B. Variation in hospital mortality associated with inpatient surgery. *N. Engl. J. Med.* **2009**, *361*, 1368–1375. https://doi.org/10.1056/NEJMsa0903048.
38. Reddy, H.G.; Shih, T.; Englesbe, M.J.; Shannon, F.L.; Theurer, P.F.; Herbert, M.A.; Paone, G.; Bell, G.F.; Prager, R.L. Analyzing "failure to rescue": Is this an opportunity for outcome improvement in cardiac surgery? *Ann. Thorac. Surg.* **2013**, *95*, 1976–1981. https://doi.org/10.1016/j.athoracsur.2013.03.027.
39. Milojevic, M.; Bond, C.; He, C.; Shannon, F.L.; Clark, M.; Theurer, P.F.; Prager, R.L. Failure to rescue: Variation in mortality after cardiac surgery. *Interact. Cardiovasc. Thorac. Surg.* **2021**, *33*, 848–856. https://doi.org/10.1093/icvts/ivab188.
40. Edwards, F.H.; Ferraris, V.A.; Kurlansky, P.A.; Lobdell, K.W.; He, X.; O'Brien, S.M.; Furnary, A.P.; Rankin, J.S.; Vassileva, C.M.; Fazzalari, F.L.; et al. Failure to Rescue Rates After Coronary Artery Bypass Grafting: An Analysis From The Society of Thoracic Surgeons Adult Cardiac Surgery Database. *Ann. Thorac. Surg.* **2016**, *102*, 458–464. https://doi.org/10.1016/j.athoracsur.2016.04.051.
41. Magouliotis, D.E.; Xanthopoulos, A.; Zotos, P.A.; Arjomandi Rad, A.; Tatsios, E.; Bareka, M.; Briasoulis, A.; Triposkiadis, F.; Skoularigis, J.; Athanasiou, T. The Emerging Role of "Failure to Rescue" as the Primary Quality Metric for Cardiovascular Surgery and Critical Care. *J. Clin. Med.* **2023**, *12*, 4876. https://doi.org/10.3390/jcm12144876.
42. Zotos, P.A.; Androutsopoulou, V.; Scarci, M.; Minervini, F.; Cioffi, U.; Xanthopoulos, A.; Athanasiou, T.; Magouliotis, D.E. Failure to Rescue and Lung Resections for Lung Cancer: Measuring Quality from the Operation Room to the Intensive Care Unit. *Cancers* **2025**, *17*, 2784. https://doi.org/10.3390/cancers17172784.

43. Sinha, S.; Ata Ozdemir, B.; Khalid, U.; Karthikesalingam, A.; Poloniecki, J.D.; Thompson, M.M.; Holt, P.J. Failure-to-rescue and interprovider comparisons after elective abdominal aortic aneurysm repair. *Br. J. Surg.* **2014**, *101*, 1541–1550. https://doi.org/10.1002/bjs.9633.
44. Silber, J.H.; Romano, P.S.; Rosen, A.K.; Wang, Y.; Even-Shoshan, O.; Volpp, K.G. Failure-to-rescue: Comparing definitions to measure quality of care. *Med. Care* **2007**, *45*, 918–925. https://doi.org/10.1097/MLR.0b013e31812e01cc.
45. Zakko, J.; Premkumar, A.; Logan, A.J.; Sneddon, J.M.; Brock, G.N.; Pawlik, T.M.; Mokadam, N.A.; Whitson, B.A.; Lampert, B.C.; Washburn, W.K.; et al. Textbook outcome: A novel metric in heart transplantation outcomes. *J. Thorac. Cardiovasc. Surg.* **2024**, *167*, 1077–1087.e13. https://doi.org/10.1016/j.jtcvs.2023.02.019.
46. Bakhtiyar, S.S.; Sakowitz, S.; Ali, K.; Coaston, T.; Verma, A.; Chervu, N.L.; Benharash, P. Textbook outcomes in heart transplantation: A quality metric for the modern era. *Surgery* **2023**, *174*, 21–29. https://doi.org/10.1016/j.surg.2023.03.014.
47. Prabhu, N.K.; Nellis, J.R.; Moya-Mendez, M.; Hoover, A.; Medina, C.; Meza, J.M.; Allareddy, V.; Andersen, N.D.; Turek, J.W. Textbook outcome for the Norwood operation-an informative quality metric in congenital heart surgery. *JTCVS Open* **2023**, *15*, 394–405. https://doi.org/10.1016/j.xjon.2023.05.003.
48. Prabhu, N.K.; Moya-Mendez, M.E.; Kang, L.; Medina, C.K.; McCrary, A.W.; Allareddy, V.; Overbey, D.; Turek, J.W. Textbook Outcome for Superior Cavopulmonary Connection: A Metric for Single Ventricle Heart Surgery. *World J. Pediatr. Congenit. Heart Surg.* **2024**, *15*, 303–312. https://doi.org/10.1177/21501351231215261.
49. Filiberto, A.C.; Banks, C.A.; Sickels, A.; Novak, Z.; Beck, A.W. Sex Differences in Textbook Outcomes among Adults Undergoing Elective Abdominal Aortic Aneurysm Repair. *J. Am. Coll. Surg.* **2025**, *240*, 552–560. https://doi.org/10.1097/XCS.0000000000001303.
50. Oden-Brunson, H.V.; Huang, Y.; Schmid, B.P.; Ferreira, Z.C.; Maximus, S.; Miranda, J.A.; Greenleaf, E.; Saqib, N.U.; Macedo, T.A.; Kalra, M.; et al. Textbook outcomes as a composite quality metric following fenestrated-branched endovascular aortic repair. *J. Vasc. Surg.* **2026**, *in press*. https://doi.org/10.1016/j.jvs.2026.08.036.
51. Hounschell, C.A.; Sanner-Dixon, K.; Broski, J.; Lambou, K.; Selby, L.V. The "Textbook Outcome" Isn't Textbook: Significant Variation in "Textbook Outcome" Definitions following a Systematic Review of the Surgical Oncology Literature. *Ann. Surg. Oncol.* **2026**, *33*, 2388–2401. https://doi.org/10.1245/s10434-025-18781-w.
52. Zhang, Y.; Wang, B.; Gao, W.; Hong, S. Textbook outcome in cancer surgery and its prognostic value: A systematic review and meta-analysis. *Eur. J. Cancer Prev.* **2026**, *35*, 534–546. https://doi.org/10.1097/CEJ.0000000000001018.
53. Gregersen, J.S.; Solstad, T.U.; Achiam, M.P.; Olsen, A.A. Textbook outcome in esophagogastric oncological surgery and long-term survival—A systematic review and meta-analysis. *Eur. J. Surg. Oncol.* **2026**, *52*, 112062. https://doi.org/10.1016/j.ejso.2026.112062.
54. Dawood, Z.S.; Khalil, M.; Waqar, U.; Banani, I.; Alidina, Z.; Pawlik, T.M. Use of textbook outcome as a quality metric in hepato-pancreaticobiliary surgery: A systematic review and meta-analysis. *J. Gastrointest. Surg.* **2025**, *29*, 102005. https://doi.org/10.1016/j.gassur.2025.102005.
55. Carbonell-Morote, S.; Daiko, H.; de Manzoni, G.; Gisbertz, S.; Yang, H.K.; Polkowski, W.; Ramia, J.M.; Pera, M. International Delphi consensus to define textbook outcome parameters for oncological gastrectomy (togs project). *World J. Surg. Oncol.* **2026**, *24*, 251. https://doi.org/10.1186/s12957-026-04314-6.
56. Siegel, J.B.; Von Ebers, R.; Philips, P.; Egger, M.E.; Scoggins, C.R.; Martin, R.C.G. Defining textbook outcome for distal pancreatectomy by systematic review. Is it clinically relevant? *Surgery* **2026**, *196*, 110202. https://doi.org/10.1016/j.surg.2026.110202.
57. Shahian, D.M.; He, X.; O'Brien, S.M.; Grover, F.L.; Jacobs, J.P.; Edwards, F.H.; Welke, K.F.; Suter, L.G.; Drye, E.; Shewan, C.M.; et al. Development of a clinical registry-based 30-day readmission measure for coronary artery bypass grafting surgery. *Circulation* **2014**, *130*, 399–409. https://doi.org/10.1161/CIRCULATIONAHA.113.007541.
58. Hannan, E.L.; Racz, M.J.; Walford, G.; Ryan, T.J.; Isom, O.W.; Bennett, E.; Jones, R.H. Predictors of readmission for complications of coronary artery bypass graft surgery. *JAMA* **2003**, *290*, 773–780. https://doi.org/10.1001/jama.290.6.773.
59. Gulliford, M.C.; Adams, G.; Ukoumunne, O.C.; Latinovic, R.; Chinn, S.; Campbell, M.J. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J. Clin. Epidemiol.* **2005**, *58*, 246–251. https://doi.org/10.1016/j.jclinepi.2004.08.012.
60. Mbekwe Yepnang, A.M.; Caille, A.; Eldridge, S.M.; Giraudeau, B. Association of intracluster correlation measures with outcome prevalence for binary outcomes in cluster randomised trials. *Stat. Methods Med. Res.* **2021**, *30*, 1988–2003. https://doi.org/10.1177/09622802211026004.
61. Adams, J.L.; Mehrotra, A.; Thomas, J.W.; McGlynn, E.A. Physician cost profiling—Reliability and risk of misclassification. *N. Engl. J. Med.* **2010**, *362*, 1014–1021. https://doi.org/10.1056/NEJMsa0906323.

62. Krell, R.W.; Hozain, A.; Kao, L.S.; Dimick, J.B. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surg.* **2014**, *149*, 467–474. https://doi.org/10.1001/jamasurg.2013.4249.
63. Merlo, J.; Chaix, B.; Ohlsson, H.; Beckman, A.; Johnell, K.; Hjerpe, P.; Råstam, L.; Larsen, K. A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J. Epidemiol. Community Health* **2006**, *60*, 290–297. https://doi.org/10.1136/jech.2004.029454.
64. Austin, P.C.; Merlo, J. Intermediate and advanced topics in multilevel logistic regression analysis. *Stat. Med.* **2017**, *36*, 3257–3277. https://doi.org/10.1002/sim.7336.
65. Krumholz, H.M.; Brindis, R.G.; Brush, J.E.; Cohen, D.J.; Epstein, A.J.; Furie, K.; Howard, G.; Peterson, E.D.; Rathore, S.S.; Smith, S.C.; et al. Standards for statistical models used for public reporting of health outcomes: An American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. *Circulation* **2006**, *113*, 456–462. https://doi.org/10.1161/CIRCULATIONAHA.105.170769.
66. Bratzler, D.W.; Normand, S.L.; Wang, Y.; O'Donnell, W.J.; Metersky, M.; Han, L.F.; Rapp, M.T.; Krumholz, H.M. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. *PLoS ONE* **2011**, *6*, e17401. https://doi.org/10.1371/journal.pone.0017401.
67. Shahian, D.M.; Bowdish, M.E.; Bloom, J.P.; Wyler von Ballmoos, M.C.; Edgerton, J.R.; Antman, M.S.; Kurlansky, P.A.; Lobdell, K.W.; Cleveland, J.C.; Gaudino, M.F.L.; et al. The Society of Thoracic Surgeons Coronary Artery Bypass Graft Composite Measure: 2021 Methodology Update. *Ann. Thorac. Surg.* **2022**, *113*, 1954–1961. https://doi.org/10.1016/j.athoracsur.2021.06.036.
68. Shahian, D.M.; Normand, S.L. What is a performance outlier? *BMJ Qual. Saf.* **2015**, *24*, 95–99. https://doi.org/10.1136/bmjqs-2015-003934.
69. Spiegelhalter, D.J. Funnel plots for comparing institutional performance. *Stat. Med.* **2005**, *24*, 1185–1202. https://doi.org/10.1002/sim.1970.
70. Shahian, D.M.; Wolf, R.E.; Iezzoni, L.I.; Kirle, L.; Normand, S.L. Variability in the measurement of hospital-wide mortality rates. *N. Engl. J. Med.* **2010**, *363*, 2530–2539. https://doi.org/10.1056/NEJMsa1006396.
71. Wong, C.W.Y.; Yu, D.S.F.; Li, P.W.C.; Chan, B.S. The prognostic impacts of frailty on clinical and patient-reported outcomes in patients undergoing coronary artery or valvular surgeries/procedures: A systematic review and meta-analysis. *Ageing Res. Rev.* **2023**, *85*, 101850. https://doi.org/10.1016/j.arr.2023.101850.
72. Collins, G.S.; Moons, K.G.M.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; Ghassemi, M.; Liu, X.; Reitsma, J.B.; van Smeden, M.; et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378. https://doi.org/10.1136/bmj-2023-078378.
73. Riley, R.D.; Ensor, J.; Snell, K.I.E.; Harrell, F.E.; Martin, G.P.; Reitsma, J.B.; Moons, K.G.M.; Collins, G.; van Smeden, M. Calculating the sample size required for developing a clinical prediction model. *BMJ* **2020**, *368*, m441. https://doi.org/10.1136/bmj.m441.
74. Steyerberg, E.W.; Vergouwe, Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur. Heart J.* **2014**, *35*, 1925–1931. https://doi.org/10.1093/eurheartj/ehu207.
75. Evans, S.R.; Rubin, D.; Follmann, D.; Pennello, G.; Huskins, W.C.; Powers, J.H.; Schoenfeld, D.; Chuang-Stein, C.; Cosgrove, S.E.; Fowler, V.G.; et al. Desirability of Outcome Ranking (DOOR) and Response Adjusted for Duration of Antibiotic Risk (RADAR). *Clin. Infect. Dis.* **2015**, *61*, 800–806. https://doi.org/10.1093/cid/civ495.
76. Myles, P.S.; Shulman, M.A.; Heritier, S.; Wallace, S.; McIlroy, D.R.; McCluskey, S.; Sillar, I.; Forbes, A. Validation of days at home as an outcome measure after surgery: A prospective cohort study in Australia. *BMJ Open* **2017**, *7*, e015828. https://doi.org/10.1136/bmjopen-2017-015828.
77. Prager, R.L.; Armenti, F.R.; Bassett, J.S.; Bell, G.F.; Drake, D.; Hanson, E.C.; Heiser, J.C.; Johnson, S.H.; Plasman, F.B.; Shannon, F.L.; et al. Cardiac surgeons and the quality movement: The Michigan experience. *Semin. Thorac. Cardiovasc. Surg.* **2009**, *21*, 20–27. https://doi.org/10.1053/j.semtcvs.2009.03.008.
78. Engelman, D.T.; Ben Ali, W.; Williams, J.B.; Perrault, L.P.; Reddy, V.S.; Arora, R.C.; Roselli, E.E.; Khoynezhad, A.; Gerdisch, M.; Levy, J.H.; et al. Guidelines for Perioperative Care in Cardiac Surgery: Enhanced Recovery After Surgery Society Recommendations. *JAMA Surg.* **2019**, *154*, 755–766. https://doi.org/10.1001/jamasurg.2019.1153.
79. Pocock, S.J.; Ariti, C.A.; Collier, T.J.; Wang, D. The win ratio: A new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur. Heart J.* **2012**, *33*, 176–182. https://doi.org/10.1093/eurheartj/ehr352.

80. Ribeiro, T.; Malhotra, A.K.; Bondzi-Simpson, A.; Eskander, A.; Ahmadi, N.; Wright, F.C.; McIsaac, D.I.; Mahar, A.; Jerath, A.; Coburn, N.; et al. Days at home after surgery as a perioperative outcome: Scoping review and recommendations for use in health services research. *Br. J. Surg.* **2024**, *111*, znae278. https://doi.org/10.1093/bjs/znae278.
81. Bell, M.; Eriksson, L.I.; Svensson, T.; Hallqvist, L.; Granath, F.; Reilly, J.; Myles, P.S. Days at Home after Surgery: An Integrated and Efficient Outcome Measure for Clinical Trials and Quality Assurance. *EClinicalMedicine* **2019**, *11*, 18–26. https://doi.org/10.1016/j.eclinm.2019.04.011.
82. Steel, R.P.; Field, J.D.; Dieleman, J.M. The value of days alive and out of hospital 30 days after surgery as an outcome measure in a major cardiac surgical center in Australia. *JTCVS Open* **2026**, *29*, 101556. https://doi.org/10.1016/j.xjon.2025.101556.
83. Dijs-Elsinga, J.; Otten, W.; Versluijs, M.M.; Smeets, H.J.; Kievit, J.; Vree, R.; van der Made, W.J.; Marang-van de Mheen, P.J. Choosing a hospital for surgery: The importance of information on quality of care. *Med. Decis. Mak.* **2010**, *30*, 544–555. https://doi.org/10.1177/0272989X09357474.
84. Abbas, A.; Dalmacy, D.; Paro, A.; Pawlik, T.M. Impact of neighborhood characteristics on textbook outcome following major surgery. *Am. J. Surg.* **2022**, *224*, 959–964. https://doi.org/10.1016/j.amjsurg.2022.04.007.
85. Boyle, L.; Curtis, E.; Paine, S.J.; Tamatea, J.; Lumley, T.; Merry, A.F. Using Days Alive and Out of Hospital to measure inequities and explore pathways through which inequities emerge after coronary artery bypass grafting in Aotearoa New Zealand: A secondary data analysis using a retrospective cohort. *BMJ Open* **2025**, *15*, e093479. https://doi.org/10.1136/bmjopen-2024-093479.
86. Shahian, D.M.; Edwards, F.H.; Jacobs, J.P.; Prager, R.L.; Normand, S.L.; Shewan, C.M.; O'Brien, S.M.; Peterson, E.D.; Grover, F.L. Public reporting of cardiac surgery performance: Part 1: History, rationale, consequences. *Ann. Thorac. Surg.* **2011**, *92*, S2–S11. https://doi.org/10.1016/j.athoracsur.2011.06.100.
87. Williamson, P.R.; Altman, D.G.; Bagley, H.; Barnes, K.L.; Blazeby, J.M.; Brookes, S.T.; Clarke, M.; Gargon, E.; Gorst, S.; Harman, N.; et al. The COMET Handbook: Version 1.0. *Trials* **2017**, *18*, 280. https://doi.org/10.1186/s13063-017-1978-4.
88. Kirkham, J.J.; Davis, K.; Altman, D.G.; Blazeby, J.M.; Clarke, M.; Tunis, S.; Williamson, P.R. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med.* **2017**, *14*, e1002447. https://doi.org/10.1371/journal.pmed.1002447.
89. Staniszewska, S.; Brett, J.; Simera, I.; Seers, K.; Mockford, C.; Goodlad, S.; Altman, D.G.; Moher, D.; Barber, R.; Denegri, S.; et al. GRIPP2 reporting checklists: Tools to improve reporting of patient and public involvement in research. *BMJ* **2017**, *358*, j3453. https://doi.org/10.1136/bmj.j3453.
90. Francis, J.; Prothasis, S.; George, J.; Stoica, S. Patient-reported outcome measures in congenital heart surgery: A systematic review. *Cardiol. Young* **2023**, *33*, 337–341. https://doi.org/10.1017/S1047951123000057.
91. Chan, P.S.; Jones, P.G.; Arnold, S.A.; Spertus, J.A. Development and validation of a short version of the Seattle angina questionnaire. *Circ. Cardiovasc. Qual. Outcomes* **2014**, *7*, 640–647. https://doi.org/10.1161/CIRCOUTCOMES.114.000967.
92. McNamara, R.L.; Spatz, E.S.; Kelley, T.A.; Stowell, C.J.; Beltrame, J.; Heidenreich, P.; Tresserras, R.; Jernberg, T.; Chua, T.; Morgan, L.; et al. Standardized Outcome Measurement for Patients With Coronary Artery Disease: Consensus From the International Consortium for Health Outcomes Measurement (ICHOM). *J. Am. Heart Assoc.* **2015**, *4*, e001767. https://doi.org/10.1161/JAHA.115.001767.
93. Herdman, M.; Gudex, C.; Lloyd, A.; Janssen, M.F.; Kind, P.; Parkin, D.; Bonsel, G.; Badia, X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual. Life Res.* **2011**, *20*, 1727–1736. https://doi.org/10.1007/s11136-011-9903-x.
94. Lansac, E.; Veen, K.M.; Joseph, A.; Blancarte Jaber, P.; Sossi, F.; Das-Gupta, Z.; Aktaa, S.; Sádaba, J.R.; Thourani, V.H.; Dahle, G.; et al. The First International Consortium for Health Outcomes Measurement (ICHOM) Standard Dataset for Reporting Outcomes in Heart Valve Disease: Moving From Device- to Patient-Centered Outcomes: Developed by a multisociety taskforce coordinated by the Heart Valve Society (HVS) including the American Heart Association (AHA), the American College of Cardiology (ACC), the European Association for Cardio-Thoracic Surgery (EACTS), the European Society of Cardiology (ESC), The Society of Thoracic Surgeons (STS), the Australian & New Zealand Society of Cardiac & Thoracic Surgeons (ANZSCTS), the International Society for Applied Cardiovascular Biology (ISACB), the International Society for Minimally Invasive Cardiothoracic Surgery (ISMICS), the South African Heart Association (SHA), Heart Valve Voice, and Global Heart Hub. *Circ. Cardiovasc. Qual. Outcomes* **2025**, *18*, e000128. https://doi.org/10.1161/HCQ.0000000000000128.
95. Spertus, J.A.; Jones, P.G. Development and Validation of a Short Version of the Kansas City Cardiomyopathy Questionnaire. *Circ. Cardiovasc. Qual. Outcomes* **2015**, *8*, 469–476. https://doi.org/10.1161/CIRCOUTCOMES.115.001958.
96. Subramanian, M.; Kozower, B.D.; Brown, L.M.; Khullar, O.V.; Fernandez, F.G. Patient-Reported Outcomes in Cardiothoracic Surgery. *Ann. Thorac. Surg.* **2019**, *107*, 294–301. https://doi.org/10.1016/j.athoracsur.2018.06.005.

97. Koch, C.G.; Li, L.; Kaplan, G.A.; Wachterman, J.; Shishehbor, M.H.; Sabik, J.; Blackstone, E.H. Socioeconomic position, not race, is linked to death after cardiac surgery. *Circ. Cardiovasc. Qual. Outcomes* **2010**, *3*, 267–276. https://doi.org/10.1161/CIRCOUTCOMES.109.880377.
98. Charles, E.J.; Mehaffey, J.H.; Hawkins, R.B.; Fonner, C.E.; Yarboro, L.T.; Quader, M.A.; Kiser, A.C.; Rich, J.B.; Speir, A.M.; Kron, I.L.; et al. Socioeconomic Distressed Communities Index Predicts Risk-Adjusted Mortality After Cardiac Surgery. *Ann. Thorac. Surg.* **2019**, *107*, 1706–1712. https://doi.org/10.1016/j.athoracsur.2018.12.022.
99. Fiscella, K.; Burstin, H.R.; Nerenz, D.R. Quality measures and sociodemographic risk factors: To adjust or not to adjust. *JAMA* **2014**, *312*, 2615–2616. https://doi.org/10.1001/jama.2014.15372.
100. Shahian, D.M.; Jacobs, J.P.; Badhwar, V.; D'Agostino, R.S.; Bavaria, J.E.; Prager, R.L. Risk Aversion and Public Reporting. Part 1: Observations From Cardiac Surgery and Interventional Cardiology. *Ann. Thorac. Surg.* **2017**, *104*, 2093–2101. https://doi.org/10.1016/j.athoracsur.2017.06.077.
101. Shahian, D.M.; Jacobs, J.P.; Badhwar, V.; D'Agostino, R.S.; Bavaria, J.E.; Prager, R.L. Risk Aversion and Public Reporting. Part 2: Mitigation Strategies. *Ann. Thorac. Surg.* **2017**, *104*, 2102–2110. https://doi.org/10.1016/j.athoracsur.2017.06.076.
102. Joynt Maddox, K.E.; Reidhead, M.; Hu, J.; Kind, A.J.H.; Zaslavsky, A.M.; Nagasako, E.M.; Nerenz, D.R. Adjusting for social risk factors impacts performance and penalties in the hospital readmissions reduction program. *Health Serv. Res.* **2019**, *54*, 327–336. https://doi.org/10.1111/1475-6773.13133.
103. Krumholz, H.M.; Bernheim, S.M. Considering the role of socioeconomic status in hospital outcomes measures. *Ann. Intern. Med.* **2014**, *161*, 833–834. https://doi.org/10.7326/M14-2308.
104. Timmermans, M.J.C.; Houterman, S.; Daeter, E.D.; Danse, P.W.; Li, W.W.; Lipsic, E.; Roefs, M.M.; van Veghel, D. Using real-world data to monitor and improve quality of care in coronary artery disease: Results from the Netherlands Heart Registration. *Neth. Heart J.* **2022**, *30*, 546–556. https://doi.org/10.1007/s12471-022-01672-0.
105. Edgerton, J.R.; Herbert, M.A.; Hamman, B.L.; Ring, W.S. Can use of an administrative database improve accuracy of hospital-reported readmission rates? *J. Thorac. Cardiovasc. Surg.* **2018**, *155*, 2043–2047. https://doi.org/10.1016/j.jtcvs.2017.11.071.
106. Hirji, S.A.; Percy, E.D.; Zogg, C.K.; Vaduganathan, M.; Kiehm, S.; Pelletier, M.; Kaneko, T. Thirty-Day Nonindex Readmissions and Clinical Outcomes After Cardiac Surgery. *Ann. Thorac. Surg.* **2020**, *110*, 484–491. https://doi.org/10.1016/j.athoracsur.2019.11.042.
107. Shawon, M.S.R.; Lujic, S.; Joshi, Y.; Jorm, L. Readmission destination following cardiac surgery and its association with mortality outcomes: A population-based retrospective study. *Lancet Reg. Health West. Pac.* **2024**, *51*, 101189. https://doi.org/10.1016/j.lanwpc.2024.101189.
108. Benedetto, U.; Dimagli, A.; Sinha, S.; Cocomello, L.; Gibbison, B.; Caputo, M.; Gaunt, T.; Lyon, M.; Holmes, C.; Angelini, G.D. Machine learning improves mortality risk prediction after cardiac surgery: Systematic review and meta-analysis. *J. Thorac. Cardiovasc. Surg.* **2022**, *163*, 2075–2087.e9. https://doi.org/10.1016/j.jtcvs.2020.07.105.
109. Hui, D.S.; Iribarne, A.; Mehaffey, J.H.; Hawkins, R.B.; Lee, R.; Lawrence, K.M.; Smith, R.L.; Takayama, H.; Schwann, T.A.; Krohn, C.; et al. The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2026 Update on Outcomes and Research. *Ann. Thorac. Surg.* **2026**, *122*, 549–559. https://doi.org/10.1016/j.athoracsur.2026.05.024.
110. Akins, C.W.; Miller, D.C.; Turina, M.I.; Kouchoukos, N.T.; Blackstone, E.H.; Grunkemeier, G.L.; Takkenberg, J.J.; David, T.E.; Butchart, E.G.; Adams, D.H.; et al. Guidelines for reporting mortality and morbidity after cardiac valve interventions. *Ann. Thorac. Surg.* **2008**, *85*, 1490–1495. https://doi.org/10.1016/j.athoracsur.2007.12.082.
111. VARC-3 Writing Committee; Généreux, P.; Piazza, N.; Alu, M.C.; Nazif, T.; Hahn, R.T.; Pibarot, P.; Bax, J.J.; Leipsic, J.A.; Blanke, P.; et al. Valve Academic Research Consortium 3: Updated endpoint definitions for aortic valve clinical research. *Eur. Heart J.* **2021**, *42*, 1825–1857. https://doi.org/10.1093/eurheartj/ehaa799.
112. Horwitz, L.I.; Grady, J.N.; Cohen, D.B.; Lin, Z.; Volpe, M.; Ngo, C.K.; Masica, A.L.; Long, T.; Wang, J.; Keenan, M.; et al. Development and validation of an algorithm to identify planned readmissions from claims data. *J. Hosp. Med.* **2015**, *10*, 670–677. https://doi.org/10.1002/jhm.2416.
113. Sterne, J.A.; White, I.R.; Carlin, J.B.; Spratt, M.; Royston, P.; Kenward, M.G.; Wood, A.M.; Carpenter, J.R. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ* **2009**, *338*, b2393. https://doi.org/10.1136/bmj.b2393.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.