CLINICAL ENDPOINT ADJUDICATION

CLINICAL ENDPOINT ADJUDICATION

Much ado about nothing?

Authors:

Dr Mohammed N. Meah (MB ChB), Dr Martin Denvir (PhD), Professor Nicholas L. Mills (PhD), Professor John Norrie (MSc), Professor David E. Newby (DSc)

British Heart Foundation Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:

Professor David Newby, Room SU.314, Chancellor's Building, University of Edinburgh, 49 Little France Crescent, Edinburgh, EH16 4SB

Tel: +44 131 242 6515

Email: d.e.newby@ed.ac.uk

Word Count: 2,000

Abstract

Event adjudication is considered to be a cornerstone of best practice for clinical trials. Scrutiny of trial endpoints by an independent committee blinded to the trial intervention, provides a robust platform and confidence for the primary trial findings. However, this apparent "gold-standard" is a complex process where data can be missing or incomplete, and decisions are rigidly protocolised whilst interpretation is subjective. Moreover, there is conflicting evidence in the literature regarding its ability to reduce misclassification of events. Although individual studies demonstrate how adjudication can add credibility and ensure that data are of sufficient quality, systematic reviews suggest the overall effect size estimates are unaltered. Indeed, the perception that endpoint decisions made by an adjudication committee are superior to that of the site investigator has never been validated. In view of the substantial financial and personnel cost that is associated with event adjudication, careful consideration should be made surrounding the methodology chosen and the benefits gained. Alternative strategies, including the use of routinely collected data, are increasingly being used to deliver randomised controlled trials more efficiently and effectively. This may ultimately be more reflective of real-world practice and indicative of healthcare impact.

In pivotal clinical effectiveness trials, the primary endpoint needs to be precisely defined and quantified because any misclassification may introduce noise and possible bias, potentially leading to incorrect trial conclusions. This is commonly addressed by having an independent clinical endpoint adjudication committee where all relevant clinical information is provided and a panel of clinical experts categorises the primary endpoint, blind to treatment allocation. This has been established as a basic cornerstone of modern robust trial methodology. However, does this apparently rigorous approach truly deliver more robust clinical findings?

The underlying premise is that adjudication will be more accurate, reduce noise and limit any misclassification by the site investigator, with any residual errors being consistently distributed across the treatment groups. However, adjudication may neither improve classification nor be an accurate reflection of 'true events'.[1] The site investigator has the advantage of full clinical context and documentation as well as an intimate understanding of the clinical situation at the site. An external clinical endpoint committee will be remote from the event and has to rely on, at times, incomplete data, poor source documentation and a formulaic protocol-driven definition of the endpoint. Site investigators can be variable in their response to requests for source data to verify events. This can be a major problem for international trials. Inevitably, differences in event classification will arise, but who is the more likely to be correct? As far as we are aware, there has yet to be a formal comparison of local assessment with central adjudication where the absolute truth is definitively known. Generally, if not exclusively, trials assume that central adjudication is more accurate than local adjudication, and so presents the central findings as the primary analysis. But is this always justified or true?

Methodology of Clinical Endpoint Committees

The effective ability to apply a standard definition decreases with increasing complexity of the clinical case and endpoint definition (**Panel: Clinical Case Example 1**). As Plesk and colleagues highlight, although definitions allow us to 'reduce and resolve' clinical cases into neat boxes, 'unpredictability and paradoxes are ever present' and some things will always remain unclear.[2] In such situations, results are subjective. As such, the methodology behind event adjudication warrants close inspection. What clinical expertise or training is required in order to sit on an adjudication panel? If central adjudicators are used, should they adjudicate every suspected outcome or only selected outcomes? Should adjudications be conducted independently or in a consensus committee? Kahan and colleagues in their detailed review of the statistical properties of various adjudication methods, felt that no single approach fits all.[3]

In most cases, adjudication panels rely on information from site investigators and this can be insensitive, conservative and restricted to deciding whether events the site has deemed worthy of reporting meet the endpoint. Although attempts can be made to screen for unreported events (**Appendix: Clinical Trial Example 1**),[4,5] they are costly and resource intensive. These problems not only apply to disease diagnoses, but also to death. Many countries have incomplete death certification and the cause of death can be very difficult to determine. Indeed, it can be based on the testimony of witnesses, colleagues or estranged family members (**Panel: Clinical Case Example 2**). In a meta-analysis of 9 clinical trials with a total of 9,259 centrally adjudicated deaths, approximately 16% had an undetermined cause.[6] There are also cultural and religious issues in attributing cause of death. For example, the acceptability of recording suicide varies considerable across the globe.[7] Clinical endpoint

adjudication committees often resort to classifying a number of events by consensus rather than unanimously, underlining that subjectivity and uncertainty remains.

What is the variability and cost of Clinical Endpoint Adjudication?

There can be marked variability in the attribution of a clinical endpoint with site investigators and clinical endpoint adjudication committees often disagreeing, especially with complex or subjective endpoints. Which opinion is most valid? One could argue that the clinical endpoint committees select only definitive cases excluding the more questionable ones. However, the evidence for this is lacking. Hallen and colleagues found that in a blinded "re-adjudication" of 10 of the most challenging cases and 10 randomly selected consistently adjudicated cases, the re-adjudicated outcomes changed in 11 cases: a discordance rate of 55%.[8] Does this mean the trial conclusions should have been reconsidered?

Adjudication is expensive.[9] Data needs to be collated, redacted to remove participant or treatment identifiable information, and sent to the trial coordinating centre. This is time consuming and often an iterative process requiring repeated communications between the sponsor, the site investigator and the clinical endpoint adjudication committee (**Appendix: Clinical Trial Example 2**).[10] The overall financial and personnel costs should not be underestimated. For international trials, this can additionally involve addressing issues of transmitting data across regulatory borders.

Is adjudication better than site reporting when determining endpoints?

Advocates for adjudication point to evidence that where there is disagreement between clinical event committees and site investigators, the participant's prognosis is often

worse.[11,12] This highlights that disagreements are almost always over endpoint definition in complex and high-risk cases. Despite this, effect size estimates often remain unchanged regardless of using adjudicated or site reported endpoints (**Appendix: Clinical Trial Examples 1 and 2**). These findings are not limited to trials where events are defined by biomarkers, such as troponin concentrations and myocardial infarction, but also where endpoints can be clearly defined, such as studies of patients with stroke.[23,24] Moreover, there are examples where site investigator reporting demonstrates greater effect sizes of more clinically relevant events than clinical endpoint adjudication which identified milder subclinical events (**Appendix: Clinical Trial Example 3**).[13]

Are these one-off examples? Systematic reviews have assessed the impact of clinical endpoint adjudication committee decisions regarding endpoint classification and compared them to those ascribed by site investigators. A meta-analysis of 10 trials by Pogue and colleagues[14] concluded that after reviewing over 95,000 patients and 9,000 events, no changes were detected in the treatment effect due to adjudication. A subsequent COCHRANE review of 47 randomised controlled trials found that treatment effect estimates did not differ, although there were differences where site investigators were unblinded to the treatment allocation.[15] It concluded that independent adjudication may be important, but raised doubts about the appropriate use of adjudication in double blind randomised controlled trials.

A common perception is that adjudication leads to confidence in the clinical trial findings. However, there are examples where this is not the case (**Appendix: Clinical Trial Example 4**).[16–20] Safety endpoints are also often seen as necessary to adjudicate but there are importance differences in the nuance of reporting and analysis here. Efficacy is usually

evaluating a single central outcome, safety analyses are looking for non-specific indications of harm across a spectrum of multiple outcomes. This added complexity may lead to important safety issues being misattributed by the clinical endpoint adjudication committee (**Appendix: Clinical Trial Example 5**).[21,22]

What are the alternatives?

Many countries around the world have unified health records data, especially where there is advanced national public healthcare provision. Some observers have been critical of the absence of endpoint adjudication in trials using these data and point to inaccuracies of hospital coding and statistics.[25] In systems where payment to the healthcare service is reliant on coding data, external factors such as reimbursement incentives or local practice variations can cause bias.[26] In large international trials, a further concern is the heterogeneity and reporting biases of different healthcare systems. Despite this, routinely collected hospital admission statistics have been used to conduct research for many years. Guidelines have helped to standardise definitions for trialists and registries.[27] This has led to the delivery of randomised controlled trials, and is gaining popularity.

A systematic review in 2001 found accuracy was high in the United Kingdom, especially in diagnostic codes.[28] In 2012, Burns *et al* conducted a further systematic review of 32 studies[29] and found that since 2002, the accuracy of hospital coding, particularly in primary diagnosis, had improved from 74 to 96%. Whilst there remains a degree of variability especially for surgical procedures, the current drive for healthcare quality improvement has also enhanced the accuracy of hospital coding further.[30,31] The literature therefore suggests that routinely collected data are robust for use in research and clinical trials.[29]

In the West of Scotland Coronary Prevention Study (WOSCOPS),[32] independently adjudicated clinical endpoint data were compared with routinely collected electronic health record data. There was excellent agreement with 100% of deaths and >95% of non-fatal clinical events being identified through health record linkage. As a result, subsequent follow up for over 20 years has now been performed entirely through routinely collected data.[34,35] Similar findings have also been found in the more contemporary aspirin for primary prevention in persons with diabetes mellitus (ASCEND) trial, where hospital episode statistics in England were compared with adjudicated clinical endpoints and effect size estimates for the primary outcome were again very similar (personal communication, Jane Armitage, University of Oxford).

Many trials are now using routinely collected health data for clinical endpoint assessments. The Scottish computed tomography (CT) of the heart (SCOT-HEART) trial[36] aimed to establish the benefit of CT coronary angiography when implemented into routine clinical practice. The clinical endpoint of coronary heart disease death or non-fatal myocardial infarction was identified through routinely collected health records data. With the introduction of CT coronary angiography, the national healthcare system observed lower rates of myocardial infarction, and this is ultimately the most important outcome for the healthcare provider. If site investigator reporting or clinical endpoint adjudication had been used, it is highly likely that absolute numbers of events would have differed, but the overall effect size is unlikely to change. This use of national electronic health record data has the potential for being automated, relatively independent of site or trial investigators, and markedly efficient.[37]

Routinely collected healthcare record data can also be used for endpoint adjudication if required. The *High-S*ensitivity *T*roponin in the *E*valuation of patients with suspected *A*cute *C*oronary *S*yndrome (*High-STEACS*) trial compared outcomes in consecutive patients evaluated using two cardiac troponin tests to determine whether implementing a more sensitive test improved diagnosis and reduced subsequent attendance with myocardial infarction or cardiovascular death.[38] All deaths and hospital attendances were adjudicated using linked routine healthcare data. Quick access and lower costs made this method an attractive alternative standard approach, although it would be limited to countries that have access to comprehensive and robust systems to capture data from electronic health records.

In the United Kingdom, the National Cancer Intelligence Network combined data from eight cancer registries, the Office for National Statistics and the Hospital Episode Statistics to create the National Cancer Data Repository (NCDR).[39] These data have again demonstrated very high levels of consistency with clinical trial data: concordance of 99% for treatment and 96% for outcomes including 100% for 30-day mortality and near identical survival at 5-years.[40] However, these approaches do require substantial upfront investment to standardise and to collate the data but once achieved, they are reliable and robust.[41]

Conclusions

Adjudication is an important tool, but like all tools, it is not appropriate for every situations (**Table**). There are important limitations which need to be considered when deciding whether to use it. The belief that a diagnosis made by endpoint adjudication is superior to the site investigator has never been substantiated. Indeed, the evidence suggests that treatment effect estimates rarely differ between site investigators and clinical endpoint committee. It is

difficult to conclude that adjudication is a gold standard that should be applied to every study. The use of routinely collected healthcare data has several strengths including being generally independent, comprehensive and highly cost effective, as well as being truly reflective of the impact on the healthcare system within which the intervention is being assessed. Whilst scientifically this may be less robust, the approach is arguably more important and relevant for the health care system and society since this will be how a new healthcare intervention will be applied in the real world. This begs the question of whether such approaches should be the gold-standard when assessing the impact of implementing healthcare interventions in the real world.

Acknowledgements

The authors are supported by the British Heart Foundation (CH/09/002, SP/12/10/29922, RG/16/10/32375, FS/16/14/32023, SP/17/12/32960, RE/18/5/34216, CS/18/4/34074, FS/19/46/34445), Wellcome Trust (WT103782AIA) and Chief Scientist Office (CZB/4/588).

Author Contributions

DEN wrote the first draft, all authors contributed to the revision and drafting of the manuscript.

Declarations of Interest

Authors are, or have been, involved in clinical trials using site investigator reported events, clinical endpoint adjudication and routinely collected healthcare data including academic and industry (pharmaceutical or device companies) led studies.

References

1. Granger CB, Vogel V, Cummings SR, et al. Do we need to adjudicate major clinical events? *Clin Trials*. 2008;5(1):56-60. doi:10.1177/1740774507087972
2. Plsek PE, Greenhalgh T. Complexity science: The challenge of complexity in health care. [Article 1 in series of 4]. *BMJ*. 2001;323(7313):625-628. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1121189/pdf/625.pdf%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/11557716.
3. Kahan BC, Feagan B, Jairath V. A comparison of approaches for adjudicating outcomes in clinical trials. *Trials*. 2017;18(1):1-14. doi:10.1186/s13063-017-1995-3
4. Mahaffey KW, Held C, Wojdyla DM, et al. Ticagrelor effects on myocardial infarction and the impact of event adjudication in the PLATO (platelet inhibition and patient outcomes) trial. *J Am Coll Cardiol*. 2014;63(15):1493-1499. doi:10.1016/j.jacc.2014.01.038
5. Wallentin L, Becker RC, Budaj A, et al. Ticagrelor versus Clopidogrel in Patients with Acute Coronary Syndromes. *N Engl J Med*. 2009;361(11):1045-1057.
6. Fanaroff AC, Clare R, Pieper KS, et al. Frequency, Regional Variation, and Predictors of Undetermined Cause of Death in Cardiometabolic Clinical Trials: A Pooled Analysis of 9259 Deaths in 9 Trials. *Circulation*. 2019;139(7):863-873. doi:10.1161/CIRCULATIONAHA.118.037202
7. Goldsmith SK, Pellmar TC, Kleinman AM, Bunney WE. *National Research Council. Reducing Suicide: A National Imperative*.; 2002.
8. Hallén J, Maggioni AP, Lopez-De-Sa E, et al. Reproducibility of in-hospital worsening heart failure event adjudication in the RELAX-AHF-EU trial. *Eur J Heart Fail*. 2019:1-2.

doi:10.1002/ejhf.1574

9. Calvo G, McMurray JJV, Granger CB, et al. Large streamlined trials in cardiovascular disease. *Eur Heart J*. 2014;35(9):544-548. doi:10.1093/eurheartj/eht535

10. Bennett-Guerrero E, Ferguson TB, Lin M, et al. Effect of an implantable gentamicin-collagen sponge on sternal wound infections following cardiac surgery: A randomized trial. *JAMA - J Am Med Assoc*. 2010;304(7):755-762. doi:10.1001/jama.2010.1152

11. Sepehrvand N, Zheng Y, Armstrong PW, et al. Alignment of site versus adjudication committee-based diagnosis with patient outcomes: Insights from the Providing Rapid out of Hospital Acute Cardiovascular Treatment 3 trial. *Clin Trials*. 2016;13(2):140-148. doi:10.1177/1740774515601437

12. Mahaffey KW, Harrington RA, Akkerhuis M, et al. Disagreements between central clinical events committee and site investigator assessments of myocardial infarction end- points in an international clinical trial: Review of the PURSUIT study. *Curr Control Trials Cardiovasc Med*. 2001;2(4):187-194. doi:10.1186/CVM-2-4-187

13. Olivier CB, Bhatt DL, Leonardi S, et al. Central Adjudication Identified Additional and Prognostically Important Myocardial Infarctions in Patients Undergoing Percutaneous Coronary Intervention. *Circ Cardiovasc Interv*. 2019;12(7):1-7. doi:10.1161/circinterventions.118.007342

14. Pogue J, Walter SD, Yusuf S. Evaluating the benefit of event adjudication of cardiovascular outcomes in large simple RCTs. *Clin Trials*. 2009;6(3):239-251. doi:10.1177/1740774509105223

15. Ndounga Diakou LA yma., Trinquart L, Hróbjartsson A, et al. Comparison of central adjudication of outcomes and onsite outcome assessment on treatment effect estimates. *Cochrane database Syst Rev*. 2016;3(3):MR000043.

doi:10.1002/14651858.MR000043.pub2

16. Stone GW, Pieter Kappetein A, Sabik JF, et al. Five-year outcomes after PCI or CABG for left main coronary disease. *N Engl J Med*. 2019;381(19):1820-1830. doi:10.1056/NEJMoa1909406

17. Mäkikallio T, Holm NR, Lindsay M, et al. Percutaneous coronary angioplasty versus coronary artery bypass grafting in treatment of unprotected left main stenosis (NOBLE): a prospective, randomised, open-label, non-inferiority trial. *Lancet*. 2016;388(10061):2743-2752. doi:10.1016/S0140-6736(16)32052-9

18. Jaffe AS. Third Universal Definition of Myocardial Infarction. *Clin Biochem*. 2013;46(1-2):1-4. doi:10.1016/j.clinbiochem.2012.10.036

19. Ruel M, Farkouh ME. Why NOBLE and EXCEL Are Consistent with Each Other and with Previous Trials. *Circulation*. 2017;135(9):822-824. doi:10.1161/CIRCULATIONAHA.116.027159

20. Ruel M, Falk V, Farkouh ME, et al. Myocardial revascularization trials: Beyond the printed word. *Circulation*. 2018;138(25):2943-2951. doi:10.1161/CIRCULATIONAHA.118.035970

21. Holmes DR, Reddy VY, Turi ZG, et al. Percutaneous closure of the left atrial appendage versus warfarin therapy for prevention of stroke in patients with atrial fibrillation: a randomised non-inferiority trial. *Lancet*. 2009;374(9689):534-542. doi:10.1016/S0140-6736(09)61343-X

22. Mandrola J, Foy A, Naccarelli G. Percutaneous left atrial appendage closure is not ready for routine clinical use. *Hear Rhythm*. 2018;15(2):298-301. doi:10.1016/j.hrthm.2017.10.007

23. Claiborne Johnston S, Donald Easton J, Farrant M, et al. Clopidogrel and aspirin in

acute ischemic stroke and high-risk TIA. *N Engl J Med*. 2018;379(3):215-225. doi:10.1056/NEJMoa1800410

24. Johnston SC, Amarenco P, Albers GW, et al. Ticagrelor versus aspirin in acute stroke or transient ischemic attack. *N Engl J Med*. 2016;375(1):35-43. doi:10.1056/NEJMoa1603060

25. Kaul S. Evaluating the Evidence for Coronary Computed Tomography Angiography as the Noninvasive Test of Choice for Patients With Stable Chest Pain. *JAMA Cardiol*. 2019;4(3):199-200. doi:10.1001/jamacardio.2018.4332

26. Seltzer JH, Heise T, Carson P, et al. Use of endpoint adjudication to improve the quality and validity of endpoint assessment for medical device development and post marketing evaluation: Rationale and best practices. A report from the cardiac safety research consortium. *Am Heart J*. 2017;190:76-85. doi:10.1016/j.ahj.2017.05.009

27. Spitzer E, McFadden E, Vranckx P, et al. Critical Appraisal of Contemporary Clinical Endpoint Definitions in Coronary Intervention Trials: A Guidance Document. *JACC Cardiovasc Interv*. 2019;12(9):805-819. doi:https://doi.org/10.1016/j.jcin.2018.12.031

28. Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. *J Public Health Med*. 2001;23(3):205-211.

29. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. *J Public Health (Bangkok)*. 2012;34(1):138-148. doi:10.1093/pubmed/fdr054

30. Mahbubani K, Georgiades F, Goh EL, et al. Clinician-directed improvement in the processes and systems. 2018;5(1):47-51.

31. Heywood NA, Gill MD, Charlwood N, Brindle R, Kirwan CC. Improving accuracy of clinical coding in surgery: collaboration is key. *J Surg Res*. 2016;204(2):490-495.

doi:10.1016/j.jss.2016.05.023

32. The West of Scotland Coronary Prevention Study Group. A coronary primary prevention study of Scottish men aged 45–64 years: Trial design. *J Clin Epidemiol*. 1992;45(8):849-860. doi:10.1016/0895-4356(92)90068-X

33. The West of Scotland Coronary Prevention Study Group. Computerised record linkage: Compared with traditional patient follow-up methods in clinical trials and illustrated in a prospective epidemiological study. *J Clin Epidemiol*. 1995;48(12):1441-1452. doi:10.1016/0895-4356(95)00530-7

34. Ford I, Murray H, McCowan C, Packard CJ. Long-term safety and efficacy of lowering low-density lipoprotein cholesterol with statin therapy 20-year follow-up of west of Scotland coronary prevention study. *Circulation*. 2016;133(11):1073-1080. doi:10.1161/CIRCULATIONAHA.115.019014

35. Kashef MA, Giugliano G. Legacy effect of statins: 20-year follow up of the West of Scotland Coronary Prevention Study (WOSCOPS). *Glob Cardiol Sci Pract*. 2017;2016(4). doi:10.21542/gcsp.2016.35

36. Newby D. CT coronary angiography in patients with suspected angina due to coronary heart disease (SCOT-HEART): An open-label, parallel-group, multicentre trial. *Lancet*. 2015;385(9985):2383-2391. doi:10.1016/S0140-6736(15)60291-4

37. Held C. When do we need clinical endpoint adjudication in clinical trials? *Ups J Med Sci*. 2019;124(1):42-45. doi:10.1080/03009734.2018.1516706

38. Shah ASV, Anand A, Strachan FE, et al. High-sensitivity troponin in the evaluation of patients with suspected acute coronary syndrome: a stepped-wedge, cluster-randomised controlled trial. *Lancet*. 2018;392(10151):919-928. doi:10.1016/S0140-6736(18)31923-8

39. Public Health England. National Cancer Intelligence Network. http://www.ncin.org.uk/home. Accessed January 15, 2020.

40. Morris EJA, Jordan C, Thomas JD, et al. Comparison of treatment and outcome information between a clinical trial and the National Cancer Data Repository. *Br J Surg*. 2011;98(2):299-307. doi:10.1002/bjs.7295

41. Kilburn LS, Aresu M, Banerji J, Barrett-Lee P, Ellis P, Bliss JM. Can routine data be used to support cancer clinical trials? A historical baseline on which to build: Retrospective linkage of data from the TACT (CRUK 01/001) breast cancer trial and the National Cancer Data Repository. *Trials*. 2017;18(1):1-9. doi:10.1186/s13063-017-2308-6

42. First International Study of Infarct Survival (ISIS-1) Collaborative Group. Randomised Trial of Intravenous Atenolol Among 16 027 Cases of Suspected Acute Myocardial Infarction: Isis-1. *Lancet*. 1986;328(8498):57-66. doi:10.1016/S0140-6736(86)91607-7

43. Newby DE, Adamson PD, Berry C, et al. Coronary CT Angiography and 5-Year Risk of Myocardial Infarction. *N Engl J Med*. 2018;379:924-933. doi:https://doi.org/10.1056/NEJMoa1805971

SUPPLEMENTARY Material

CLINICAL ENDPOINT ADJUDICATION
Much ado about nothing?

Authors:

Dr Mohammed N. Meah (MB ChB), Dr Martin Denvir (PhD), Professor Nicholas L. Mills (PhD), Professor John Norrie (MSc), Professor David E. Newby (DSc)

British Heart Foundation Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, United Kingdom

Clinical Trial Example 1

The platelet inhibition and patient outcomes (PLATO) trial[5] took a novel approach to screen for unreported clinical endpoint events. In addition to data supplied by site investigators, the clinical endpoint committee had access to background biomarker analysis performed on serial samples taken from study participants. This background screening led to 101 more events being adjudicated as true myocardial infarctions than reported by the site investigators.[4] Without such background screening, the true prevalence of events will almost always be underestimated and this further reduces the sensitivity of identifying potentially important clinical events. It is interesting to note that, whilst the point estimate of the treatment effect size remained unchanged whether this approach was incorporated or not, the confidence interval around the estimate was broader using site investigator reported events. Here the additional screening eliminated the reliance on site reporting and therefore led to a more statistically robust conclusion.[4]

Clinical Trial Example 2

A study looking at the effect of implantable gentamicin-collagen sponge on sternal wound infections had an adjudication panel that included three experts in infectious diseases who were blinded to the treatment assignment.[9] Possible infections were identified by triggered events on an electronic case report form which led to a review of the entire medical records of 128 patient at considerable cost. The investigators presented the data from both the principal investigators' and the adjudication panel, and found results were similar with no benefit in the gentamicin arm as compared to the control arm.

Clinical Trial Example 3

In the cangrelor versus standard therapy to achieve optimal management of platelet inhibition (CHAMPION PHOENIX) trial, the endpoint of myocardial infarction following percutaneous coronary intervention varied over 3-fold between the clinical endpoint committee and the site investigators.[12] It is perhaps unsurprising that site investigators were more reluctant to attribute peri-procedural events as myocardial infarction. In cases where there may be bias, or a likelihood of poor interobserver reproducibility, adjudication helps to prevent high rates of misclassifications. Despite this, the primary endpoint findings remained valid and indeed the events identified by the site investigators (also blind to treatment allocation) demonstrated the largest treatment effect size, perhaps underlining that they report the more clinically significant events. Although the number of endpoints can vary substantially, the 'misclassification' of events did not lead to a difference in study conclusions.

	Number of Myocardial Infarctions Reported	Adjusted Odds Ratio (95% Confidence Interval)	P-Value
Adjudication Committee	462	5·35 (2·56-11·20)	<0·001
Site Investigator	143	9.08 (4·01-20·50)	<0·001

Clinical Trial Example 4

In the 5-year outcomes in the evaluation of drug-eluting stents versus coronary artery bypass surgery for left main revascularisation (EXCEL) trial, coronary stenting was reported as non-inferior to surgery.[15] The trial endpoints were all adjudicated but there is controversy over the way in which myocardial infarction was defined and adjudicated. Unlike previous trials of left main stem revascularisation which found stenting to be inferior to surgery,[16] the primary composite outcome of death, myocardial infarction or stroke in the EXCEL trial included periprocedural myocardial infarction. The definition of periprocedural myocardial infarction in EXCEL differed from the Universal Definition,[17] favoured the use of creatine kinase-MB over troponin, applied the same thresholds for surgery and stenting, and did not require ancillary evidence from coronary angiography or cardiac imaging.[18] As a consequence, the endpoint committee identified an excess of procedural events of uncertain clinical significance in those undergoing surgery that was offset by an increase in spontaneous myocardial infarctions over time in those undergoing stenting.[19] Given all-cause mortality was higher in patients undergoing stenting, and the secondary outcome of myocardial infarction defined by the Universal Definition criteria remains unreported, doubts remain as to whether these strategies are truly equivalent. Indeed, the European Association of Cardio-Thoracic Surgeons has since made the unprecedented decision to reverse its endorsement of the recommendations based on the EXCEL trial in the joint myocardial revascularisation guidelines.

Clinical Trial Example 5

In the percutaneous closure of the left atrial appendage versus warfarin therapy for prevention of stroke in patients with atrial fibrillation (PROTECT-AF) trial, excessive bleeding was a key safety endpoint.[21] Adjudication was independent but unblinded, and the authors reported a 90% reduction in rates of haemorrhagic stroke but this has been subsequently challenged.[22] A Food and Drug Administration review found uneven adjudication of haemorrhagic stroke in the trial whereby falls with subsequent subdural haematomas were labelled as a positive event in the warfarin group (5 of the total 10), but not in the device group (3 in total). This led to an overestimate of safety benefit of the treatment intervention due to endpoint adjudication.

Table. Methods of clinical endpoint ascertainment in clinical trials

	Advantages	Disadvantages	When to use?	Case Example
End-point adjudication	- Improves specificity by consistently applying a standardised definition.[4,5] - Limits ascertainment bias particularly in unblinded trials Improved scientific acceptability of trial results	- Reliant on events reported by site investigators - Data may be incomplete or inconsistent.[21,22] - Substantial cost and effort. [10] - Inflexible application of endpoint diagnosis may lead to false conclusions	- Regulatory drug trials - Endpoint is difficult to define - Investigators or participants are unblinded to trial intervention (e.g. device-based trials)	PLATO [5]
Site investigator reporting	- Decision based on full clinical context - Improved sensitivity - Potentially as effective at determining treatment effect	- May be subject to selection bias from both over and underreporting. - Reduced specificity due to inconsistent application of endpoint diagnoses	- Endpoints are clearly defined and easily standardised - Where site investigators have expertise and training	CRASH studies [48–50]
Routinely collected data	- High efficiency - Reflects impact of intervention on the healthcare system - Ceding control of endpoint definitions gives added level of independence. - Potential use with artificial intelligence and machine learning techniques.	- Reliant on comprehensive and robust data capture from healthcare information systems. - Dependent on a stable population. - Definitions of data collected may change during the study	- Assessment of intervention on the healthcare system.[40,41] - Long-term follow-up - When costs need to be limited	SCOT-HEART [43]

Panel

Clinical Case Example 1.

A participant in an investigator-led multicentre randomised controlled trial (>1500 participants in >30 centres) dies in hospital 3 days after being admitted with a fall at home. Excellent hospital records are available to allow the clinical endpoint committee to adjudicate the cause of death. The patient has a background of aortic stenosis, coronary heart disease, frailty, heart failure and a recent diagnosis of breast cancer. On arrival in hospital, the patient has evidence of widespread trauma including multiple fractured ribs, bilateral haemo-thoraces, liver contusions and pelvic fluid on CT scan. The ECG shows 1-2 mm of anterior ST elevation and cardiac troponin is mildly elevated. There is evidence of heart failure with marked oedema and lung crepitations on auscultation. The patient is made not for resuscitation and dies within 3 days of admission. The local investigator certifies the death as due to 1 A) Heart failure, 1 B) acute coronary syndrome and 2) aortic stenosis. The adjudication committee are equally split on the cause of death as cardiovascular and non-cardiovascular. The local investigators opinion is considered important by the endpoint committee and the death is designated as “cardiovascular” despite severe trauma being the apparent direct cause of death.

Clinical Case Example 2.

A participant of a major respiratory, pharma-sponsored, international, multicentre randomised controlled trial (>15,000 participants in over 1,200 centres in more than 40 countries) is lost to follow up. A private detective is hired to find the participant. The detective attends the participants home at a trailer park but cannot contact them. A passer-by walking their dog shares with the detective that the participant had died. The circumstances of their death are unknown, and the death was not registered (not a legal requirement in their place of residence). The deceased participant had no savings and was cremated by their fellow residents with an impromptu pyre and wake at the trailer park. The fatal primary endpoint was therefore defined by the testimony of the passer-by walking his dog