

RESEARCH

Open Access



# Accuracy of healthcare systems data for identifying cardiovascular outcomes after stroke due to intracerebral haemorrhage in the United Kingdom

Alice Hosking<sup>1</sup>, Jacqueline Stephen<sup>2</sup>, Jonathan Drever<sup>1</sup>, William N. Whiteley<sup>1,4</sup>, Cathie L. M. Sudlow<sup>2,4</sup>, Rustam Al-Shahi Salman<sup>1,3\*</sup> and for the RESTART Trial Steering Committee

## Abstract

**Background** Healthcare systems data (HCSD) could improve the efficiency of clinical trials, but their accuracy and validity are uncertain. Our objective was to assess the accuracy of HCSD as the sole method of outcome detection in the REstart or STop Antithrombotics Randomised Trial (RESTART; ISRCTN71907627) compared with adjudicated questionnaire follow-up and compare estimates of treatment effect.

**Methods** RESTART was a prospective, open, assessor-blind, parallel-group randomised controlled trial (RCT) of antiplatelet therapy after intracerebral haemorrhage (ICH) in the UK.

We included 496 (92%) of 537 RESTART participants, who were resident in England or Scotland at randomisation. Computerised randomisation incorporating minimisation allocated participants (1:1) to start or avoid antiplatelet therapy.

RESTART used annual questionnaires to detect its primary outcome (recurrent ICH) and secondary outcome (a composite of haemorrhagic or ischemic major adverse cardiovascular events [MACE]) over a median of 2.0 years; an independent adjudication committee verified outcomes using medical records and brain imaging. We obtained ICD10-coded HCSD on hospital admissions and deaths in England and Scotland to identify primary and secondary outcomes. We compared HCSD with a reference standard of adjudicated outcomes. We estimated the effects of antiplatelet therapy using HCSD alone in a Cox proportional hazards model adjusted for minimisation variables.

**Results** In the original RESTART trial, 31 people experienced a primary outcome event. HCSD had sensitivity of 84% (95% CI 66 to 95%) and positive predictive value of 68% (51 to 82%) for recurrent ICH. HCSD estimated an effect of antiplatelet therapy (adjusted hazard ratio [aHR] 0.51, 95% CI 0.27 to 0.98;  $p=0.044$ ) that was almost identical to adjudicated outcomes (aHR 0.51, 95% CI 0.25 to 1.03;  $p=0.060$ ). HCSD had sensitivity of 84% (76 to 91%) and positive predictive value of 78% (69 to 85%) for MACE, on which HCSD estimated an effect of antiplatelet therapy (aHR 0.81, 95% CI 0.56 to 1.16;  $p=0.247$ ) that was similar to adjudicated outcomes (aHR 0.65, 95% CI 0.44 to 0.95;  $p=0.025$ ).

\*Correspondence:

Rustam Al-Shahi Salman

[Rustam.al-shahi@ed.ac.uk](mailto:Rustam.al-shahi@ed.ac.uk)

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Conclusions** In a RCT of antiplatelet therapy for people with ICH, HCSD was reasonably accurate and provided similar estimates of treatment effect compared with adjudicated outcomes.

**Trial registration** [ISRCTN71907627](https://www.isrctn.com/ISRCTN71907627). Registered on 25 April 2013.

**Keyword** Healthcare systems data; Intracerebral haemorrhage; Trial outcome adjudication

## Background

Over the last half-century, carefully conducted, robust randomised controlled trials (RCTs) have informed the secondary prevention of cardiovascular diseases in everyday clinical practice. In stroke medicine, RCTs have shown the benefits of antihypertensive therapy [1], statin therapy [2], carotid endarterectomy [3], antiplatelet therapy [4], and dual antiplatelet therapy [5]. For intracerebral haemorrhage (ICH), RCTs have provided evidence of the benefits of acute [6] and long-term blood pressure management [7], and intensive care bundles [8]. However, RCTs are expensive, time-consuming, and carbon-intensive to conduct because of their staffing and resource requirements. These resource requirements include the burden of face-to-face, postal, or remote methods of follow-up for outcomes and their adjudication, which also squeeze RCT budgets and limit the duration of follow-up [9, 10].

Healthcare systems data (HCSD) from electronic health records, administrative databases, or disease registers can be used to design and streamline participant identification, recruitment, consent, randomisation, follow-up, and detect clinical outcome events in RCTs. HCSD has the potential to improve the convenience and reduce the cost of RCTs for cardiovascular disease and for stroke specifically [11, 12]. The proportion of RCTs in the UK using HCSD for detection of primary or secondary outcomes appears to be increasing since 2019 [13]. To be confident in RCTs' results, outcome data need to be as accurate, complete, and unbiased as possible, resulting in valid and reliable estimates of treatments' effects. For RCTs using HCSD to identify outcomes, utility assessments comparing HCSD to RCTs' alternative primary methods of outcome detection should inform whether HCSD meet these requirements, but few such assessments exist in the UK or elsewhere [14, 15]. A systematic review of ten RCTs of cardiovascular event prevention found good agreement between HCSD and adjudicated outcomes for death and some cardiovascular outcomes, with similar directions and magnitudes of treatment effects [16]. In individual RCTs comparing HCSD with adjudicated outcomes, one RCT of aspirin for primary prevention of cardiovascular events had similar findings including very similar effects of aspirin on major bleeding [17]. However, the completeness of HCSD was

insufficient for detecting cardiovascular outcomes in a prostate cancer RCT [18].

There is uncertainty about the utility of HCSD for ascertaining cardiovascular outcomes and determining treatment effects in people with stroke and especially people with multiple cardiovascular diseases. This population is particularly challenging because of the occurrence of both ischemic and haemorrhagic major adverse cardiovascular events (MACE). These can be difficult to distinguish from non-specific International Classification of Diseases 10th Revision (ICD-10) codes, can recur within the same hospital admission, can result in multiple hospital admissions for the same event, or can be attributed as the primary cause of death in someone who dies from the consequences of disability long after a MACE.

Therefore, we sought to compare the accuracy and treatment effect estimates of HCSD (as the sole method of follow-up) versus adjudicated cardiovascular outcomes detected by questionnaires in a prospectively planned study within the Restart or Stop Antithrombotics Randomised Trial (RESTART; [ISRCTN71907627](https://www.isrctn.com/ISRCTN71907627)) of long-term antiplatelet therapy after stroke due to intracerebral haemorrhage (ICH), which was conducted in the UK between 2013 and 2018 [19].

## Methods

### Study design and participants

RESTART was a prospective, open, assessor-blinded, parallel-group RCT that included 537 participants at 122 hospitals in the UK (England, Scotland, Wales and Northern Ireland) between May 22, 2013, through May 31, 2018. Participants were aged  $\geq 18$  years and had taken antithrombotic therapy for the prevention of occlusive vascular disease when they developed spontaneous (non-traumatic) symptomatic ICH, discontinued antithrombotic therapy, and survived for 24 h. Participants were randomised to either start or avoid antiplatelet therapy. The protocol pre-specified a primary outcome of recurrent symptomatic ICH [20, 21]. The protocol also pre-specified a composite secondary outcome of all MACE defined by the Antithrombotic Trialists' Collaboration (nonfatal myocardial infarction, nonfatal stroke [ischemic, haemorrhagic, or uncertain cause], or death from a vascular/unknown cause) [22].

The funding application and trial protocol pre-specified that HCSD would be collected alongside other methods to identify outcome events, for which participants gave consent.

### Test methods

In RESTART, outcome events were ascertained using structured postal or telephone questionnaires completed annually by each participant (or their representative) and their primary care doctor, and from spontaneous reporting by trial sites. All outcome events underwent independent clinical adjudication by two consultant neurologists, who were blinded to treatment allocation and use of antithrombotic therapy, using all available source documentation from medical records and brain imaging. Participants' follow-up lasted between 6 months and 5 years until November 30, 2018 (median 2.0 years, interquartile range [IQR] 1.0–3.0) in all participants bar one who withdrew before the first follow-up [19, 23].

In this study, the index test for outcome events was HCSD. After trial recruitment and follow-up were complete, we requested HCSD from administrative databases of ICD-10-coded hospital admissions and deaths in England (via NHS England's Data Access Request Service) and Scotland (via Public Health Scotland's electronic Data Research and Innovation Service), but not Northern Ireland (where they were not available for research) or Wales (due to administrative issues), thereby including

the two largest UK nations where 92% of the RESTART cohort resided when they were recruited. English hospital admissions are recorded in Hospital Episode Statistics – Admitted Patient Care (HES-APC), with up to twenty diagnoses recorded as ICD-10 codes for each admission by administrators based on medical documentation at hospital discharge. Scottish hospital admissions are recorded in Scottish Morbidity Records (SMR01), with a 'main condition' and up to five 'other conditions' recorded as ICD-10 codes for each admission by administrators based on medical documentation at hospital discharge. Medical certificates of causes of death in England and Scotland involve doctors recording the disease or condition leading directly to death (part 1a), other diseases or conditions leading to 1a including the underlying cause of death, and other significant conditions contributing to death but not related to the disease or condition causing (part 2); these records are coded using ICD-10 and held by NHS Digital in England and National Records of Scotland.

We requested data between each participant's randomisation date and November 30, 2018. We mapped recurrent ICH and components of the composite outcome of all MACE (defined by the Antithrombotic Trialists' Collaboration) to ICD-10 codes (Table 1), with reference to the CALIBER chronological map of human health [24], the Antithrombotic Trialists [25] and 2002 meta-analyses [25, 26], UK Biobank definitions of stroke

**Table 1** ICD-10 codes used to identify outcomes in healthcare systems data

Outcome	ICD-10 code(s) used to identify outcomes in HCSD
<i>Primary outcome</i>	
Intracerebral haemorrhage	I61.x
<i>Secondary outcome</i>	
<i>MACE composite of all of the following:</i>	
Myocardial infarction (ST elevation or non-ST-elevation)	I21.x, I22.x
Ischemic stroke	I63.x
Symptomatic stroke of uncertain sub-type	I63.9, I64.x, G46.x
Symptomatic recurrent intracerebral haemorrhage <sup>a</sup>	I61.x
Symptomatic spontaneous intracranial haemorrhage	I60.x, I62.x, I69.0, I69.2
Symptomatic major extracranial haemorrhage <sup>a</sup>	N83.7, O71.7, O90.2, T81.0, H35.6, H43.1, H45.0, K92.0, K92.1, K92.2
Mesenteric ischemia <sup>a</sup>	K55.02, K55.04, K55.06
Symptomatic peripheral arterial occlusion <sup>a</sup>	I73.1, I73.8, I73.9, I74.3, I74.4, I74.5
Symptomatic deep vein thrombosis <sup>a</sup>	I82.4, I82.6
Symptomatic pulmonary embolism <sup>a</sup>	I26.x
Sudden cardiac death	I47.2, I49.0, I46.0, I46.9, I47.0
Other cardiovascular death	I20.x to I25.x, I30.x to I52.x, I60.x to I69.x, I70.x to I79.x, I80.x to I89.x
Deaths of undetermined cause	R99

HCSD, healthcare systems data; ICD-10, 10th version of the International Classification of Diseases

I61.x means I61 and all subsidiary codes

<sup>a</sup> Only fatal events were included in the composite MACE outcome to adhere to the Antithrombotic Trialists' collaboration definition

and myocardial infarction [27, 28], and record linkage studies of ICH [29, 30]. We searched for these pre-specified codes in fields for the disease or condition leading directly to death or the first/main diagnostic code for each hospital admission.

In this study, the reference standard was the best-available outcome data. Any hospital admissions or deaths detected by HCSD that did not match an outcome event or serious adverse event known to RESTART to within  $\pm 1$  day were independently adjudicated with source documentation to identify any outcome events that were missed by RESTART. This information was combined with the RESTART adjudicated outcomes to give the reference standard.

### Analysis

We did not perform a sample size calculation but used the largest dataset available to us by including all RESTART participants who had consented to data linkage and were resident in England or Scotland when randomised. We restricted analyses to the first recurrent ICH or MACE outcome event during follow-up, using ICD-10 codes in the primary diagnostic position in HCSD.

### Accuracy of HCSD

We restricted analyses to hospital admissions and deaths following the date a participant was discharged from hospital after randomisation in RESTART (to avoid identifying the qualifying ICH as an outcome event) and before the date of their last questionnaire follow-up. Measures of diagnostic accuracy used the following definitions. True positives were outcome events identified by both the index test (HCSD) and the reference standard (adjudicated questionnaire follow-up and site reporting in RESTART combined with adjudicated HCSD). False positives were outcomes identified by the index test, but not

by the reference standard. False negatives were outcomes identified by the reference standard but not by the index test. True negatives (i.e. people without an outcome in either dataset) could not be identified. We calculated the sensitivity and positive predictive value (PPV) of HCSD with exact 95% confidence intervals (CI).

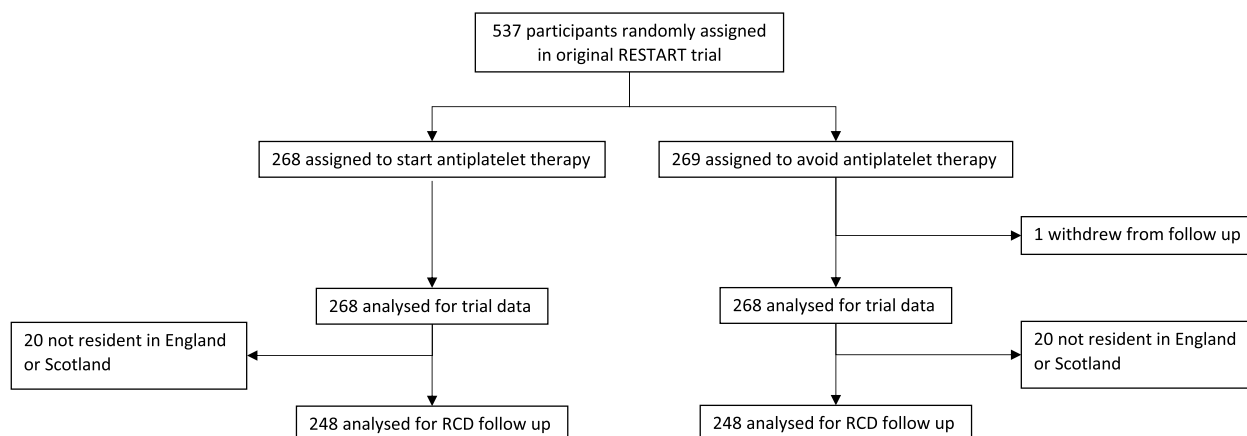
### Treatment effect estimates using HCSD

We estimated the effect of starting versus avoiding antiplatelet therapy for the included participants according to the treatment strategy assigned at randomisation using outcome events identified by HCSD alone (without reference to RESTART's adjudicated outcomes) until November 30, 2018 (allowing longer follow-up than the date of each participant's last completed questionnaire). We estimated the survival in each treatment group using a Kaplan–Meier survival analysis of time to first occurrence of an outcome event during all available follow-up after randomisation, censored at death unrelated to an outcome event or last available follow-up. We assessed the proportional hazards assumption graphically and including a treatment by log(time) interaction, and if it was met, we constructed a Cox proportional hazards regression model adjusted for all five covariates included in the minimisation algorithm to calculate an adjusted hazard ratios (aHR) as was done in RESTART [19].

## Results

### Participants and outcomes

The baseline characteristics and prognosis of participants in RESTART have been reported elsewhere [19]. Of 537 RESTART participants, one withdrew from follow-up, 29 were resident in Wales and 11 in Northern Ireland, leaving 496 (92%) who were resident in England and Scotland at randomisation to be included in these analyses (Fig. 1).



**Fig. 1** RESTART participants suitable for follow-up via healthcare systems data in England and Scotland

**Accuracy of HCSD**

Amongst the 496 participants, HCSD identified 64 recurrent ICH and 194 MACE events from ICD-10 codes, of which 51 and 168 respectively were in the primary diagnostic position, and 38 and 104 respectively were first events that occurred between hospital discharge following randomisation and before each participant's last follow-up. RESTART identified 31 first instances of recurrent ICH during trial follow-up, whereas HCSD identified 38 events in the same period. There were 26 true positives, 5 false negatives, and 12 false positives, giving a sensitivity of 84% (95% CI 66 to 95%) and PPV of 68% (51 to 82%) (Table 2). HCSD did not identify new primary outcomes that had been missed by adjudicated questionnaires. RESTART identified 96 first instances of MACE during trial follow-up, whereas HCSD identified 104 events in the same period. There were 81 true positives, 15 false negatives, and 23 false positives, giving a sensitivity of 84% (95% CI 76 to 91%) and PPV of 78% (95% CI 69 to 85%) (Table 2). HCSD identified three new MACE outcomes that had been missed by adjudicated questionnaires.

**Treatment effect estimates using HCSD**

Amongst the 496 participants, HCSD identified 41 recurrent ICH and 114 MACE events that occurred between hospital discharge following randomisation and the end of the trial on November 30, 2018. In RESTART, 35 first recurrent ICH and 110 MACE occurred during follow-up of 536 participants. The direction and magnitude of the effect of antiplatelet therapy on recurrent ICH was identical in RESTART (aHR 0.51, 95% CI 0.25–1.03;

$p=0.060$ ) and HCSD (HR=0.51, 95% CI 0.27–0.98;  $p=0.044$ ), although HCSD was more precise and identified a statistically significant effect (Fig. 2). The direction and magnitude of the effect of antiplatelet therapy on MACE was similar in RESTART (aHR 0.65, 95% CI 0.44–0.95;  $p=0.025$ ) and HCSD (HR=0.81, 95% CI 0.56–1.16;  $p=0.247$ ), although the magnitude of the point estimate of effect was somewhat attenuated in HCSD (Fig. 3).

**Discussion**

In prospectively planned analyses of HCSD conducted after the results of RESTART were known, we found that HCSD had reasonable diagnostic accuracy when compared with adjudicated outcomes. HCSD alone gave an identical estimate of treatment effect on the primary outcome of recurrent ICH (which became statistically significant) and a similar estimate of treatment effect on the secondary outcome of MACE. This suggests that biases in the ascertainment of events were similar for the two methods.

Our study has strengths: we included participants from two UK nations in the only RCT of antiplatelet therapy after ICH. HCSD performed well despite the population being multimorbid with multiple cardiovascular diseases. Whereas previous studies in this area have compared HCSD to the trial outcome [31], we undertook a thorough process of outcome adjudication and were able to compare HCSD to a best-available outcome, to give a truer estimate of the sensitivity and specificity of HCSD.

The weaknesses of our study are that we were not able to analyse the HCSD for the full original trial population of our relatively small trial, due to administrative

**Table 2** First outcomes during follow-up after randomisation in healthcare systems data compared to adjudicated outcomes

		Healthcare systems data			
		ICD-10 outcome code	No ICD-10 outcome code	Sensitivity (95% CI)	PPV (95% CI)
Adjudicated outcomes detected using annual questionnaires	Recurrent ICH				
	Outcome occurred	26 <sup>a</sup>	5 <sup>b</sup>	84% (66 to 95%)	68% (51 to 82%)
	Outcome did not occur	12 <sup>c</sup>	x		
	MACE				
	Outcome occurred	81 <sup>d</sup>	15 <sup>e</sup>	84% (76 to 91%)	78% (69 to 85%)
	Outcome did not occur	23 <sup>f</sup>	x		

PPV, positive predictive value

<sup>a</sup> 1 admission matched the second recurrent ICH, but not the first (which occurred outside the UK)

<sup>b</sup> 2 missed due to coding (I60.9 and I62.9), 2 had no recorded hospital admission, and 1 occurred during a hospital admission for another outcome

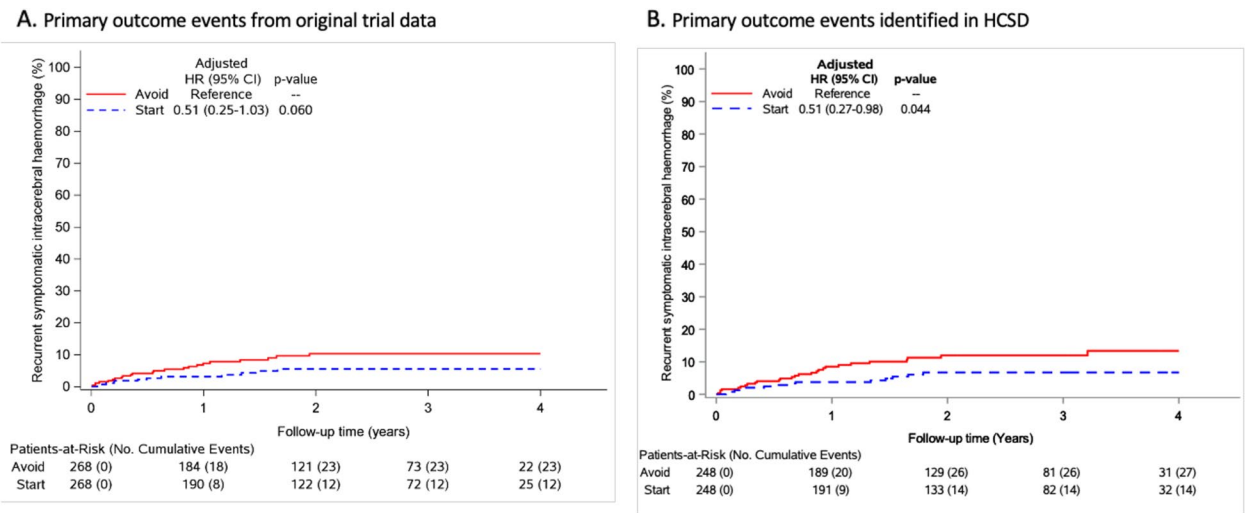
<sup>c</sup> 8 hospital admissions/deaths were not due to recurrent ICH and 4 were MACE outcomes but not recurrent ICH

<sup>d</sup> 3 admissions identified outcomes that were missed by RESTART but are included in the reference standard

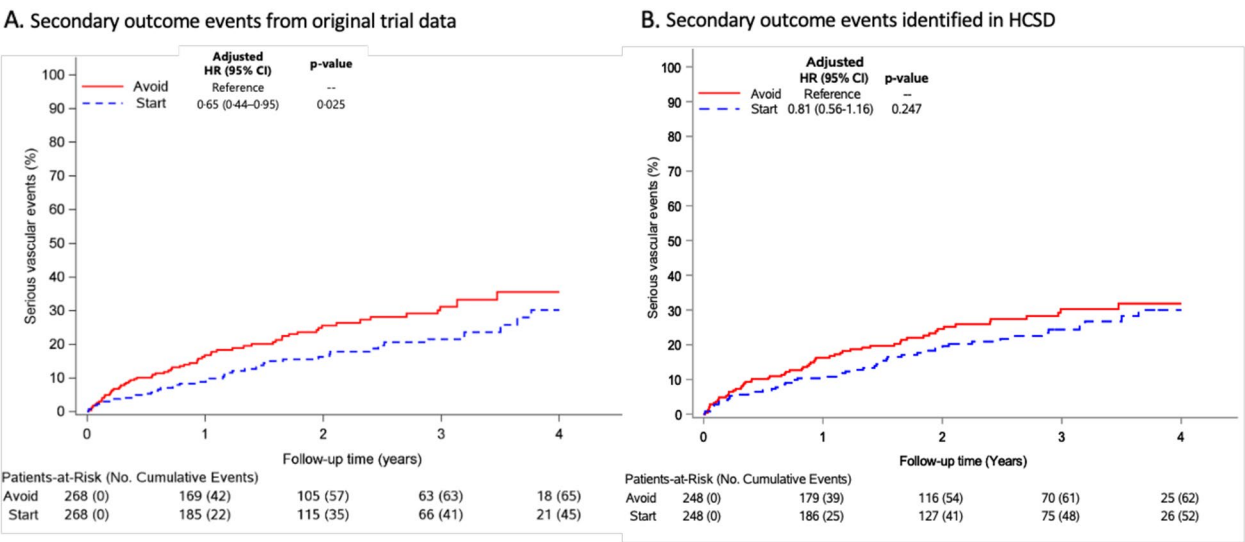
<sup>e</sup> 7 missed due to coding (I24.9, I60.9, I67.9, I25.1, I50.0, H53.8, and J18.9), 7 had no recorded hospital admission, and 1 occurred during a hospital admission for another outcome

<sup>f</sup> 11 hospital admissions/deaths were not due to MACE and 12 were outcomes but not MACE





**Fig. 2** Comparison of the effects of antiplatelet therapy on RESTART’s primary outcome of recurrent ICH using adjudicated outcomes detected by annual questionnaires for 536 participants with follow-up in the UK (A) [19] or outcomes identified using healthcare systems data for 496 participants resident in England or Scotland (B)



**Fig. 3** Comparison of the effects of antiplatelet therapy on RESTART’s secondary outcome of major adverse cardiovascular events (MACE, as defined by the Antithrombotic Trialists’ Collaboration [22] using adjudicated outcomes detected by annual questionnaires for 536 participants with follow-up in the UK (A) [19] or outcomes identified using healthcare systems data for 496 participants resident in England or Scotland (B)

and governance issues, so we did not link data for the 41 people living in Northern Ireland and Wales. The study design is susceptible to incorporation bias, which occurs when the reference standard incorporates the test under study. Our reference standard included HCSD with the adjudicated trial outcome, but this did not apply to the primary outcome and only affected the secondary outcome (we included three MACE outcomes identified by HCSD in the reference standard). Analyses may have been affected by unquantified migration of patients between and from UK nations, which would lead to an underestimate of the accuracy of HCSD, but we expect migration rates to be low in relatively elderly population. There is no agreement about how accurate HCSD need to be for use in RCT follow-up, so our values for PPV and sensitivity must be interpreted in the context of similarities in treatment effects. Our reference standard is imperfect, because traditional trial outcome ascertainment with adjudication has limitations. It is reliant on reporting of events by site investigators; source documentation can be incomplete or inconsistent; and adjudicators often

disagree, underlining the inherent subjectivity in the process [32].

Our results are relevant to MACE in a secondary prevention RCT after ICH. There may be condition-specific issues with HCSD which reduce the generalizability of our results. In observational studies, definitions used for MACE in HCSD vary, prohibiting comparison, aggregation, and replication of findings [33]. Further work is needed to comprehensively understand the accuracy of coding of cardiovascular disease outcomes in modern healthcare systems. Countries and regions have different systems and incentives for recording healthcare data which may bias results [34], and our results only apply to England and Scotland.

Previous studies have shown that the result of a clinical trial obtained through HCSD may be closer to null than when obtained through traditional follow-up methods, implying that larger sample sizes might be needed if planning follow-up with HCSD [16]. More recently, a comparison of major bleeding events recorded as adjudicated outcomes in the ASCEND trial compared to HCSD found no clinically important differences in the relative risk with treatment of events depending on data source used [35]. The result we obtained for RESTART using HCSD for the primary outcome had an identical hazard ratio to the original trial, and for secondary outcomes, the hazard ratio was similar. The accuracy of HCSD is limited by correct diagnostic coding in the data. It may be that improvements in electronic healthcare records over the past decade have increased the accuracy of HCSD. The accuracy of recording of outcome events in HCSD for a study of bladder cancer in England improved substantially between 2011 and 2017, with remuneration policies likely driving the improvement in data quality [36]. Many of the studies of HCSD in cardiovascular disease were conducted years ago, such as the West of Scotland Coronary Prevention Study, which originally began recruiting in 1995 [31].

In the current study, HCSD identified both false positive and false negative events, compared to the best-available data. The relatively large number of false positives reduced the PPV of HCSD. Our experience suggests that re-admissions can often attract incorrect coding, leading to false positives. Future work could focus on how to identify these events without sacrificing completeness. Two primary outcome events and seven secondary outcome events were false negatives due to miscoding. Access to alternative sources of information, such as national audit data or primary care data, would allow triangulation of outcomes and potentially increase the accuracy of HCSD.

Our sample size was too small to explore whether HCSD missed mild events that did not result in hospital

admission or death, but were detected by RESTART. These events could be detected with access to primary care records, but linking RCT data to primary care records is currently not possible in England due to data governance and data ownership constraints [37]. Widening the scope of access could improve the accuracy of HCSD for all clinical research.

HCSD is being used in RCTs to enhance data collection, as in the RECOVERY platform trial in COVID-19 [38], for very long-term follow-up [39], and as exclusive outcome follow-up [40, 41]. Although digitally enabled RCTs seem to offer benefits, researchers in the Bladder-Path RCT of bladder cancer treatment pathways found rapid access to outcomes from digital sources was ‘too cumbersome and expensive’ in England for incorporation into the RCT design. The same authors were able to use HCSD for one-off long-term follow-up of another RCT [42]. The MRC Clinical Trials Unit has described the protracted process of applications for HCSD from NHS Digital and the National Institute for Cardiovascular Outcomes Research, which in the case of the PATCH trial took several years [43].

In this study, we have demonstrated that HCSD has good accuracy for identifying cardiovascular outcome events and effect estimates that are comparable to adjudicated outcome events. A Delphi consensus has prioritised the remaining challenges for use of HCSD in RCTs [44]. Further work is needed to confirm the accuracy of HCSD for cardiovascular research in other healthcare systems and to enable other types of informative HCSD to be efficiently incorporated into RCTs.

## Conclusions

In a RCT of antiplatelet therapy for people with ICH, HCSD was reasonably accurate and provided similar estimates of treatment effect compared with adjudicated outcomes. Trialists could consider using healthcare systems data to detect major adverse cardiovascular events after intracerebral haemorrhage

## Abbreviations

aHR	Adjusted hazard ratio
HCSD	Healthcare systems data
ICD-10	International Classification of Diseases 10th Revision
ICH	Intracerebral haemorrhage
MACE	Major adverse cardiovascular events
MRC	Medical Research Council
PPV	Positive predictive value
RCT	Randomised controlled trial
RESTART	REstart or STop Antithrombotics Randomised Trial

## Acknowledgements

We are grateful to the RESTART participants and the Safe Haven provided by the Edinburgh International Data Facility for these analyses done in The University of Edinburgh (<https://edinburgh-international-data-facility.ed.ac.uk/security-governance/safe-haven>). The University of Edinburgh and the Lothian Health Board are co-sponsors.

Independent trial steering committee members: Colin Baigent (chairperson; University of Oxford, Oxford), Daniel Lasserson (University of Birmingham, Birmingham), Frank Sullivan (University of St Andrews, St Andrews); Johanna Carrie (patient—public representative; Edinburgh). Other trial steering committee members: Rustam Al-Shahi Salman, Martin S Dennis, Gordon D Murray, David E Newby, Peter AG Sandercock, Cathie LM Sudlow, William N Whiteley, Nikola Sprigg, David J Werring, Phil M White.

### Role of funders

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

### Authors' contributions

RA-SS (chief investigator) conceived the idea for RESTART and obtained funding and developed the protocol. RA-SS, JD, and JS designed and implemented the study. JD checked and processed the HCSD. JS was the unmasked trial statistician who did the data analyses, which were checked by AH and RA-SS. AH wrote the first draft of the manuscript, which was revised by RA-SS and reviewed by all authors who approved the final version.

### Funding

This work was funded by a British Heart Foundation Special Project grant (SP/12/2/20422).

### Data availability

A fully anonymised version of the original RESTART trial dataset used for analysis with individual participant data and a data dictionary is available for researchers to apply to use, via <https://datashare.is.ed.ac.uk/handle/10283/3265>. Written proposals will be assessed by members of the RESTART trial steering committee and a decision made about the appropriateness of the use of data. Healthcare systems data are stored in national safe havens and were accessed for one-off use and so are not available for sharing.

### Declarations

#### Ethics approval and consent to participate

The Scotland A research ethics committee approved RESTART before recruitment began (reference 12/SS/0138). The funding application and trial protocol pre-specified that HCSD would be collected alongside other methods to identify outcome events, for which participants gave consent.

#### Consent for publication

Not applicable—no identifying images or other personal or clinical details of participants are presented.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK. <sup>2</sup>The Usher Institute, The University of Edinburgh, Edinburgh, UK. <sup>3</sup>Edinburgh Clinical Trials Unit, The Usher Institute, The University of Edinburgh, Edinburgh, UK. <sup>4</sup>British Heart Foundation Data Science Centre, Health Data Research UK, London, UK.

Received: 5 August 2024 Accepted: 11 November 2024

Published online: 16 November 2024

### References

- MacMahon S, Neal B, Tzourio C, et al. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. *Lancet*. 2001;358(9287). [https://doi.org/10.1016/S0140-6736\(01\)06178-5](https://doi.org/10.1016/S0140-6736(01)06178-5).
- Amarenco P, Bogousslavsky J, Callahan A 3rd, et al. High-dose atorvastatin after stroke or transient ischemic attack. *N Engl J Med*. 2006;355(6):549–59. <https://doi.org/10.1056/NEJMoa061894>.
- Rothwell PM, Eliasziw M, Gutnikov SA, Warlow CP, Barnett HJM. Endarterectomy for symptomatic carotid stenosis in relation to clinical subgroups and timing of surgery. *Lancet*. 2004;363(9413):915–24. [https://doi.org/10.1016/S0140-6736\(04\)15785-1](https://doi.org/10.1016/S0140-6736(04)15785-1).
- Rothwell PM, Algra A, Chen Z, Diener HC, Norrving B, Mehta Z. Effects of aspirin on risk and severity of early recurrent stroke after transient ischaemic attack and ischaemic stroke: time-course analysis of randomised trials. *Lancet*. 2016;388(10042). [https://doi.org/10.1016/S0140-6736\(16\)30468-8](https://doi.org/10.1016/S0140-6736(16)30468-8).
- Bhatia K, Jain V, Aggarwal D, et al. Dual antiplatelet therapy versus aspirin in patients with stroke or transient ischemic attack: meta-analysis of randomized controlled trials. *Stroke*. 2021;52(6). <https://doi.org/10.1161/STROKEAHA.120.033033>.
- Anderson CS, Heeley E, Huang Y, et al. Rapid blood-pressure lowering in patients with acute intracerebral hemorrhage. *N Engl J Med*. 2013;368(25):2355–65. [https://doi.org/10.1056/NEJMOA1214609/SUPPL\\_FILE/NEJMOA1214609\\_DISCLOSURES.PDF](https://doi.org/10.1056/NEJMOA1214609/SUPPL_FILE/NEJMOA1214609_DISCLOSURES.PDF).
- Chapman N, Huxley R, Anderson C, et al. Effects of a perindopril-based blood pressure-lowering regimen on the risk of recurrent stroke according to stroke subtype and medical history. *Stroke*. 2004;35(1):116–21. <https://doi.org/10.1161/01.STR.0000106480.76217.6F>.
- Ma L, Hu X, Song L, et al. The third Intensive Care Bundle with Blood Pressure Reduction in Acute Cerebral Haemorrhage Trial (INTERACT3): an international, stepped wedge cluster randomised controlled trial. *Lancet*. 2023;0(0). [https://doi.org/10.1016/S0140-6736\(23\)00806-1/ATTACHMENT/ED5CEEE4-C75A-419E-B197-10F3F9A45403/MMC1.PDF](https://doi.org/10.1016/S0140-6736(23)00806-1/ATTACHMENT/ED5CEEE4-C75A-419E-B197-10F3F9A45403/MMC1.PDF).
- Martin L, Hutchens M, Hawkins C, Radnov A. How much do clinical trials cost? *Nat Rev Drug Discov*. 2017;16(6). <https://doi.org/10.1038/nrd.2017.70>.
- Adshead F, Al-Shahi Salman R, Aumonier S, et al. A strategy to reduce the carbon footprint of clinical trials. *The Lancet*. 2021;398(10297):281–2. [https://doi.org/10.1016/S0140-6736\(21\)01384-2](https://doi.org/10.1016/S0140-6736(21)01384-2).
- Khan MS, Usman MS, Talha KM, et al. Leveraging electronic health records to streamline the conduct of cardiovascular clinical trials. *Eur Heart J*. 2023;44(21):1890–909. <https://doi.org/10.1093/EURHEARTJ/EHAD171>.
- Berwanger O, MacHline-Carrion MJ. Digital health-enabled clinical trials in stroke: ready for prime time? *Stroke*. 2022;53(9):2967–75. <https://doi.org/10.1161/STROKEAHA.122.037378>.
- Toader AM, Gamble CL, Dodd S, Williamson PR. The use of healthcare systems data for RCTs. *Trials*. 2024;25(1). <https://doi.org/10.1186/S13063-023-07846-4>.
- Mc Cord KA, Al-Shahi Salman R, Treweek S, et al. Routinely collected data for randomized trials: promises, barriers, and implications. *Trials*. 2018;19(1). <https://doi.org/10.1186/s13063-017-2394-5>.
- Sydes MR, Murray ML, Ahmed S, et al. Getting our ducks in a row: the need for data utility comparisons of healthcare systems data for clinical trials. *Contemp Clin Trials*. Published online March 2024:107514. <https://doi.org/10.1016/J.CCT.2024.107514>.
- Rodrigues C, Odutayo A, Patel S, et al. Accuracy of cardiovascular trial outcome ascertainment and treatment effect estimates from routine health data: a systematic review and meta-analysis. *Circ Cardiovasc Qual Outcomes*. 2021;14(5):E007903. <https://doi.org/10.1161/CIRCOUTCOMES.120.007903>.
- Harper C, Mafham M, Herrington W, et al. Reliability of major bleeding events in UK routine data versus clinical trial adjudicated follow-up data. *Heart*. 2023;109(19):1467–72. <https://doi.org/10.1136/HEARTJNL-2023-322616>.
- Macnair A, Nankivell M, Murray ML, et al. Healthcare systems data in the context of clinical trials - a comparison of cardiovascular data from a clinical trial dataset with routinely collected data. *Contemp Clin Trials*. 2023;128. <https://doi.org/10.1016/J.CCT.2023.107162>.
- Al-Shahi Salman R, Dennis MS, Sandercock PAG, et al. Effects of antiplatelet therapy after stroke due to intracerebral haemorrhage (RESTART): a randomised, open-label trial. *The Lancet*. 2019;393(10191):2613–23. [https://doi.org/10.1016/S0140-6736\(19\)30840-2](https://doi.org/10.1016/S0140-6736(19)30840-2).
- Al-Shahi Salman R, Murray GD, Dennis MS, et al. The REstart or STOP Antithrombotics Randomised Trial (RESTART) after stroke due to intracerebral haemorrhage: statistical analysis plan for a randomised controlled trial. *Trials*. 2019;20(1). <https://doi.org/10.1186/S13063-019-3270-2>.
- Al-Shahi Salman R, Dennis MS, Murray GD, et al. The REstart or STOP Antithrombotics Randomised Trial (RESTART) after stroke due to



- intracerebral haemorrhage: study protocol for a randomised controlled trial. *Trials*. 2018;19(1). <https://doi.org/10.1186/S13063-018-2542-6>.
22. Collins R, Peto R, Hennekens C, et al. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. *The Lancet*. 2009;373(9678):1849–60. [https://doi.org/10.1016/S0140-6736\(09\)60503-1](https://doi.org/10.1016/S0140-6736(09)60503-1).
  23. Al-Shahi Salman R, Dennis MS, Sandercock PAG, et al. Effects of antiplatelet therapy after stroke caused by intracerebral hemorrhage: extended follow-up of the RESTART randomized clinical trial. *JAMA Neurol*. 2021;78(10). <https://doi.org/10.1001/jamaneurol.2021.2956>.
  24. Kuan V, Denaxas S, Gonzalez-Izquierdo A, et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit Health*. 2019;1(2):e63–77. [https://doi.org/10.1016/S25589-7500\(19\)30012-3](https://doi.org/10.1016/S25589-7500(19)30012-3).
  25. Collaboration AT. Collaborative overview of randomised trials of antiplatelet therapy: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ*. 1994;308(6921):81–106. <https://doi.org/10.1136/BMJ.308.6921.81>.
  26. Baigent C, Sudlow C, Collins R, Peto R. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ*. 2002;324(7329):71–86. <https://doi.org/10.1136/BMJ.324.7329.71>.
  27. Schnier C, Bush K, Nolan J, Sudlow C. Definitions of stroke for UK biobank phase 1 outcomes adjudication on behalf of UK Biobank Outcome Adjudication Group. Published online 2017.
  28. Schnier C, Bush K, Nolan J, Sudlow C. Definitions of acute myocardial infarction and main myocardial infarction pathological types UK biobank phase 1 outcomes adjudication on behalf of UK Biobank Outcome Adjudication Group. Published online 2017.
  29. McGuire AJ, Raikou M, Whittle I, Christensen MC. Long-term mortality, morbidity and hospital care following intracerebral hemorrhage: an 11-year cohort study. *Cerebrovasc Dis*. 2007;23(2–3):221–8. <https://doi.org/10.1159/000097645>.
  30. Schmidt LB, Goertz S, Wohlfahrt J, Melbye M, Munch TN. Recurrent intracerebral hemorrhage: associations with comorbidities and medicine with antithrombotic effects. *PLoS One*. 2016;11(11). <https://doi.org/10.1371/JOURNAL.PONE.0166223>.
  31. Barry SJE, Dinnett E, Kean S, Gaw A, Ford I. Are routinely collected NHS administrative records suitable for endpoint identification in clinical trials? Evidence from the West of Scotland Coronary Prevention Study. *PLoS ONE*. 2013;8(9): e75379. <https://doi.org/10.1371/JOURNAL.PONE.0075379>.
  32. Meah MN, Denvir MA, Mills NL, Norrie J, Newby DE. Clinical endpoint adjudication. *The Lancet*. 2020;395(10240):1878–82. [https://doi.org/10.1016/S0140-6736\(20\)30635-8](https://doi.org/10.1016/S0140-6736(20)30635-8).
  33. Bosco E, Hsueh L, McConeghy KW, Gravenstein S, Saade E. Major adverse cardiovascular event definitions used in observational analysis of administrative databases: a systematic review. *BMC Med Res Methodol*. 2021;21(1). <https://doi.org/10.1186/s12874-021-01440-5>.
  34. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*. 2018;361:1479. <https://doi.org/10.1136/BMJ.K1479>.
  35. Harper C, Maffham M, Herrington W, et al. Reliability of major bleeding events in UK routine data versus clinical trial adjudicated follow-up data. *Heart*. Published online. 2023. <https://doi.org/10.1136/heartjnl-2023-322616>.
  36. Mintz HP, Dosanjh A, Parsons HM, et al. Development and validation of a follow-up methodology for a randomised controlled trial, utilising routine clinical data as an alternative to traditional designs: a pilot study to assess the feasibility of use for the BladderPath trial. *Pilot Feasibility Stud*. 2020;6(1). <https://doi.org/10.1186/s40814-020-00713-y>.
  37. Wood A, Denholm R, Hollings S, et al. Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource. *BMJ*. 2021;373. <https://doi.org/10.1136/BMJ.N826>.
  38. Dexamethasone in hospitalized patients with COVID-19. *N Engl J Med*. 2021;384(8):693–704. [https://doi.org/10.1056/NEJMOA2021436/SUPPL\\_FILE/NEJMOA2021436\\_DATA-SHARING.PDF](https://doi.org/10.1056/NEJMOA2021436/SUPPL_FILE/NEJMOA2021436_DATA-SHARING.PDF).
  39. Whiteley WN, Gupta AK, Godec T, et al. Long-term incidence of stroke and dementia in ASCOT. *Stroke*. 2021;52(10):3088–96. <https://doi.org/10.1161/STROKEAHA.120.033489>.
  40. Green BB, Wang CY, Anderson ML, et al. An automated intervention with stepped increases in support to increase uptake of colorectal cancer screening: a randomized trial. *Ann Intern Med*. 2013;158(5 Pt 1):301–11. <https://doi.org/10.7326/0003-4819-158-5-201303050-00002>.
  41. Dregan A, Van Staa TP, McDermott L, et al. Point-of-care cluster randomized trial in stroke secondary prevention using electronic health records. *Stroke*. 2014;45(7):2066–71. <https://doi.org/10.1161/STROKEAHA.114.005713>.
  42. Mintz HP, Dosanjh ARS, Parsons H, et al. Making administrative healthcare systems clinical data the future of clinical trials: lessons from BladderPath. *BMJ Oncol*. 2023;2(1). <https://doi.org/10.1136/bmjonc-2023-000038>.
  43. Macnair A, Love SB, Murray ML, et al. Accessing routinely collected health data to improve clinical trials: recent experience of access. *Trials*. 2021;22(1). <https://doi.org/10.1186/s13063-021-05295-5>.
  44. Williams ADN, Davies G, Farrin AJ, et al. A DELPHI study priority setting the remaining challenges for the use of routinely collected data in trials: COMORANT-UK. *Trials*. 2023;24(1):1–8. <https://doi.org/10.1186/S13063-023-07251-X/TABLES/3>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.