Imperial College London

The Use of Routinely Collected Data for Interventional Research in Secondary Care:

A feasibility evaluation of a multi-centre randomised controlled trial and systematic review of current practice

Hanaya Raad

Research Associate, Imperial Clinical Trials Unit, Imperial College London.

Guy Martin

Honorary Clinical Research Fellow, Department of Surgery & Cancer, Faculty of Medicine, Imperial College London.

Victoria Cornelius

Senior Lecturer in Medical Statistics, Deputy Head of Statistics, Imperial Clinical Trials Unit, Imperial College London.

Word Count

Abstract

Background

Randomised controlled trials provide high quality evidence for the effectiveness of health service and medical interventions. The costs associated with conducting a trial is high and a substantial proportion of this cost is for prospective data collection. As a result the use of routinely collected data obtained from electronic patients' records for use in trials is receiving much attention by researchers and funders for the potential cost saving that can be offered.

This project consists of two parts. In the first part we undertook a feasibility study to test the collection, extraction, collation and cleaning processes of data obtained from Trust electronic health records (EHR). In the second part we undertook a systematic review to identify and critically appraise current experiences for UK interventional studies using Trust EHR data.

Method

The feasibility study was performed at Imperial College Healthcare NHS Trust, which comprises of five hospitals. All data was to be extracted and anonymised by NHS Trust IT staff members who would then transfer via secure transfer for storage on Imperial College secure servers for data analysis.

The systematic review included interventional studies performed in the UK evaluating interventions that collected primary outcome data from Trust EHR data. Data was extracted on study characteristics, methods for EHR data extraction and investigators experiences.

Results

In the feasibility study we found that close working relationship between the clinicians, statisticians and Trust IT allowed valuable opportunity to obtain a shared understanding of the data requirements. No trust data was extracted and transferred to the study team during the one year funded period for this grant.

The systematic review identified seven trials published since 2015 that utilised trust EHR data. They included a median of 518 participants and two sites per study. None of the studies explicitly reported poor experiences with data extraction, linkage of records, cleaning or delays in obtaining data in the discussion

Conclusion

The feasibility study demonstrated that the use of routinely collected data can be vulnerable to decisions be out of the control of the clinical investigator. While there are external factors that impact prospective individually collected data, governance hurdles and other external factors faced when using routinely collected data can be intractable. The use of routinely collected trust data hold great potential for use in interventional trials and there is evidence for success in published trials. The research community would benefit from investigators sharing their experiences in the practicalities of running such studies.

Introduction

Randomised controlled trials (RCTs) provide high quality evidence for the effectiveness of health services and medical interventions. The costs associated with conducting a trial is high and includes expensive prospective data collection. Standard practice is that bespoke case report forms are designed and a database is developed and validated for each trial. Research funders are increasing looking to maximise output whilst minimising costs. One way to achieve this is through encouraging efficiency in trial design and conduct that does not impact the quality of the research. As a result, the use of routinely collected data in electronic patients' records is receiving much attention by funders for the potential cost saving that can be offered.

Electronic health records (EHRs) in primary care have been used for many years for observational research. More recently the UK Clinical Practice Research Datalink (CPRD) has been successfully employed as a way to recruit and obtain outcome data for randomised trials in primary care. Whist these trials have shown it is feasible to use primary care electronic health record data for RCTs, the trials involved faced significant barriers in setup, conduct and analysis. [van Staa.2012, van Staa.2014 Julious 2016]

CPRD is a well-recognised source of routinely collected data for primary care research, less is known about the use of secondary care EHRs for randomised controlled trials. A feasibility study to examine the use of EHRs in five district general hospitals in the UK in 2003 demonstrated the potential of this valuable resource but highlighted limitations of access issues, variability in data structure and coding. [Williams 2003]

This project had two aims. The first was to explore the benefits, limitations and feasibility of using Trust hospital data for a multi-centre trial in planning. The second aim was to review the experiences of researchers using routinely collected hospital data to evaluate an intervention.

Objectives

The project was in two parts. In the first part we aimed to test out the collection, extraction, collation and cleaning processes of data obtained from electronic health records in Imperial NHS Trust. This served as a feasibility study to a definitive trial

that is planned to evaluate a mobile application (Streams) developed by DeepMind. [https://deepmind.com/] Streams enables clinicians to view patient data at the point of care with the aim to reduce the time taken to view key diagnostic results and ultimately improve patient outcomes. The definitive trial is planned to take place across all five hospitals in Imperial College Healthcare NHS Trust (ICHNT).

The aims of the feasibility study were to assess the following:

- Ease of data collection for routinely collected trust data and mobile app data
- Timeliness of data extraction
- Data quality, consistency, missing information and logical checks
- Success of linking hospital data to app data

In the second part we undertook a systematic review to identify and critically appraise current experiences for UK studies that evaluated an intervention using hospital EHRs.

The aims of the systematic review were to:

- To assess the number and nature of completed interventional trials in the UK using hospital EHR to collect primary outcome data
- To collate researchers' experiences of the use of these data

Methods: Feasibility Study

The feasibility study was performed at Imperial College Healthcare NHS Trust (ICHNT), which comprise of five hospitals in North West London. These are St Mary's Hospital, Charing Cross Hospital, Hammersmith Hospital, Queen Charlotte's and Chelsea Hospital and The Western Eye Hospital. The aim was to obtain daily data from each site for a four-week period. All planned baseline and outcome measures were to utilise routinely collected clinical data. Performance data and trial data was to be accessible through the Imperial College Healthcare NHS Trust IT (ICHNT IT) systems and Cerner EHR. All data collected was to be at individual level, extracted and anonymised by Imperial College Healthcare NHS Trust IT staff members who would then transfer via secure transfer system to researchers where it would be stored on Imperial College secure servers. As this was a feasibility study that included greater

than 100,000 participants across five hospitals no sample size rationale was explored as the sample size was deemed large enough to estimate feasibility parameters with good precision. The planned analysis methods were primarily descriptive estimating the following outcomes: proportion of missing data by variable; number of duplicated records; consistency of data e.g. blood test results published before the blood test results were reviewed.

Governance

The Streams project at ICHNT - including this evaluation - is overseen by a dedicated Project Board chaired by the Deputy Medical Director and reporting directly to the Trust Board. All relevant reviews and project approvals were obtained from the Imperial College Joint Research Compliance Office and ICHNT Quality Improvement, ICT and Information Governance Teams.

Results: Feasibility Study

An initial data schema for the ICHNT IT staff was developed in October 2017. (appendix 1) This provided the basis for the IT staff member to develop the code to extract data from Cernar. The trust investigator, project statisticians and ICHNT IT staff met on several subsequent occasions to review the code and preliminary data extracts for a test sample at one site. Issues identified included deleting patients with missing data and coding errors. The discussion between the clinician, statistician and IT allowed valuable opportunity to obtain a shared understanding of the data requirements and project requirements. The IT staff member developed code to extract the data and we were reliant on approval for the project to start in order to progress. Over the course of the remaining months of funding there were delays in the project start date until a final decision was made to put this project on hold. This was outside of the control of the investigators and data extracts for the feasibility trial were never obtained. (appendix 2)

This feasibility study demonstrated that the use of routinely collected data can be vulnerable to decisions be out of the control of the clinical investigator. While there are also external factors that may also impact individually collected data, the governance hurdles faced when using routinely collected data can be intractable. We were unable to meet our original project aims for this study in the limited time-funded window.

Methods: Systematic review

The inclusion and exclusion criteria for this review were developed using the recommended PICOS framework. The population of interest in this review were UK secondary healthcare in- or outpatients. This review focused on prospectively planned interventional trials undertaken in a secondary or tertiary healthcare setting. Interventions could be either medical, or aiming to improve quality of services. Any comparator was considered. The outcomes collected in this systematic review were trial characteristics (size, # sites, and population), mechanism to identify participants, primary and secondary outcomes, databases used and method of extraction. In addition we sought descriptions of researchers' experiences using EHRs in the trial. Publications of interventional studies or registered trial protocols were included. Retrospective and observational studies were excluded. Conference abstracts, methodological papers and reviews were excluded but reference lists were screened to identify for additional further studies.

The search strategy was developed through a scoping review to analyse the text words, index terms and synonyms contained in titles and abstracts for concepts used in the research question. The search strategy can be seen in appendix 3. Reference lists of included studies were also screened. Protocols were identified from searching UK Clinical Trials Gateway and WHO International Clinical Trials Registry Platform. The WHO ICTRP is an international initiative that collates trials from registries around the world including the United States's ClinicalTrials.gov, the EU Clinical Trials Register (EU-CTR) and the United Kingdoms' ISRCTN registry. The UK Clinical Trials Registry website was used as we found it included more detailed descriptions in its trials records.

Duplicates were removed and screening was in two stages. First, titles and abstracts were screened and then full texts were retrieved. One reviewer conducted the search

and screening stages and this process was repeated to check relevant studies were not missed. Data from all included studies were collated into a data extraction table. (appendix 4)

The data extraction table included general study information on authorship, type of intervention and outcomes and detailed information about the process of obtaining and extracting routinely collected data, including who, how and what was included in data extraction.

Data analysis was descriptive and a PRISMA flow diagram was created to track the number of studies retrieved, included and excluded, based on the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement. [Moher 2009] Reasons for excluding studies are also stated.

Results: Systematic review

The electronic search of Medline and EMBASE yielded 178 studies. On WHO International Clinical Trials Registry Platform (WHO ICTRP) and UK Clinical Trials Gateway (UKCTG), 607 and 179 trials records were found respectively. (Table 1) After eligibility assessment there were seven full-text articles included in the final review. The PRISMA flow diagram showing the screening and eligibility with reasons can be seen in Figure 1.

Table 1: Number of retrieved studies from database and trials registry websites.

Source	Number of retrieved studies/trials
Databases Search (Medline and EMBASE)	178
WHO ICTRP	607
UK Clinical Trials Gateway	179

All of the seven trials identified had been published since 2015. The median size of the trial were 518 participants (range from 80 to 11010 participants), with a median of two sites per trial (range 1 to 92 sites). The study characteristics of the included trial can be seen in Table 2.

Three studies identified the participants in the trial using a manual approach, three studies used a record based approach, and one study did not report how they identified participants. All studies reported the source of the data, but only four of the studies reported how the data was extract and these descriptions lacked information on frequency and timing. In five studies it was the study team that were responsible for the extraction, and in two studies data extraction was performed by the clinical or IT team. (Table 3) None of the studies explicitly reported poor experiences with data extraction, linkage of records, cleaning or delays in obtaining data. This suggests that either investigators did face any significant issues with these aspects of working with routinely collected data, which we think is unlikely or they did not report them. (Table 4)

Conclusion

We were limited in our ability to assess the feasibility of using routinely EHR Trust data in an interventional study within Imperial NHS Trust as we did not obtain final anonymised data extracts during the study period. This demonstrates that reliance on obtaining EHR data can be outside the investigators control. We found developing a close working relationship with IT NHS trust staff is a vital component to ensure correct preparation of the data for extract and to obtain the correct format. This communication and close review ensured that participants with missing data were not deleted and coding errors were identified.

The systematic review identified that some investigators have successfully preformed interventional trials utilising EHR to obtain primary outcome data but these are few and recent. It was not clear if the study authors faced no issues or did not share the issues of working with routinely collected data.

Routinely collected trust data hold great potential for use in interventional trials and is early in adoption. The research community would benefit from investigators already experienced in their field to share their experiences more widely in running such studies.

Table 2: Study characteristics of included trials in the systematic review

Title	Author	Year	**Journal**	**Size of trial (n)**	**number of sites/hospitals**	**Eligible population**	**Intervention**	**Primary Outcome**
Randomised controlled trial of GP-led in-hospital management of homeless people ('Pathway')	Hewett	2016	Clinical Medicine	410	2	Inpatient homeless adults	"Pathway" - GP enhanced care for homeless people	Cumulative duration of hospital stay (time between admission and discharge summed across all admissions within 90 days of initial admission, censored at 90 days)
'Seizure First Aid Training' for people with epilepsy who attend emergency departments, and their family and friends: study protocol for intervention development and a pilot randomised controlled trial	Noble	2015	BMJ Open	80	3	Patient with epilepsy	Self-management intervention to improve confidence and ability to manage seizure	Emergency department use in the 12 months following randomisation
A brief psychological intervention to reduce repetition of self-harm in patients admitted to hospital following a suicide attempt: a randomised controlled trial	O'Connor	2017	The Lancet Psychiatry	518	1	Patients with self-reported history of self-harm	Violatoin helpsheet - brief psychological intervention	Outcome of appointment (attended, missed, cancelled by the participant, cancelled by the service)
Effect of financial incentives on breastfeeding a cluster randomized clinical trial.	Relton, C.	2018	JAMA Pediatrics	10010	92	Breastfeeding mother-infant dyads	Financial incentive for breastfeeding	Electoralward area-level 6- to 8-week breastfeeding period prevalence, as assessed by clinicians at the routine 6- to 8-week postnatal check visit.

Use of Coronary Computed Tomographic Angiography to Guide Management of Patients with Coronary Disease.	Williams	2016	Journal American College of Cardiology	4146	1	Patients attending a rapid access chest pain clinic	Computed Tomography (CT) scanning	Rates of death, myocardial infarction (MI), and cerebrovascular disease
Self-Harm Intervention: Family Therapy (SHIFT), a study protocol for a randomised controlled trial of family therapy vers	Wright-Hughes	2015	Trials	832	40	Young people with history of self-harm	Family therapy	Rates of repetition of self-harm leading to hospital attendance 18 months after randomisation
Effectiveness of automated appointment reminders in psychosis community services: a randomised controlled trial	Kravariti	2017	British Journal of Psychiatry Open	95	1	Mental health patients with appointments at a Psychosis Community Service	Text message reminders	Outcome of appointment (attended, missed, cancelled by the participant, cancelled by the service)

Table 3: Data item extraction for each study

Author	What additional data is being collected routinely?	Where did the data come from? What database/s were used?	How data was extracted	Who did data extraction of EHR	Are patients identified from records or manually? Or mix of both.	How participants were identfied from EHR?
Hewett	Secondary and long-term outcomes	Hospital data system	Data was obtained by regular downloads.	The trial team	manual	Hospital ward staff notified the homelessness nurse of all admitted patients (18 or older) who were homeless
Noble	Identifying patients, baseline data and long-term follow up	Routinely collected hospital data Hospital Episode Statistics system	Not mentioned	Trial statistician and researchers	Records	Emergency departments generated lists of potentially eligible participants from their electronic attendance records for patients with presentation/discharge code
O'Connor	Identifying patients, secondary and follow-up outcomes	Information Services Division of the National Health Service (NHS ISD) and patient medical records	The Division of NHS ISD maintains database of hospital records and mortality data. The outcomes were extracted both from this database and from patient medical records.	Information Services Division of the National Health Service (NHS ISD) staff and from patient medical records by research staff.	records	Data on hospital re-admission were obtained from NHS ISD > Data on emergency departments re-admission were obtained from medical note > Data on self-reported self-harm history were obtaiened from admitted patients
Relton	Secondary outcomes	Routinely collected electoral ward area-leve; breastfeeding data	Not mentioned	By those delivering routine infant feeding services (midwives, health visitors, and primary care physicians) and collated by the local National Health Service Trust, Local Authority, or Child Health Information team.	records	Participants were women who opted into the scheme at electoral ward areas that are the geographic unit for which routine aggregated data on infant feeding is routinely collected

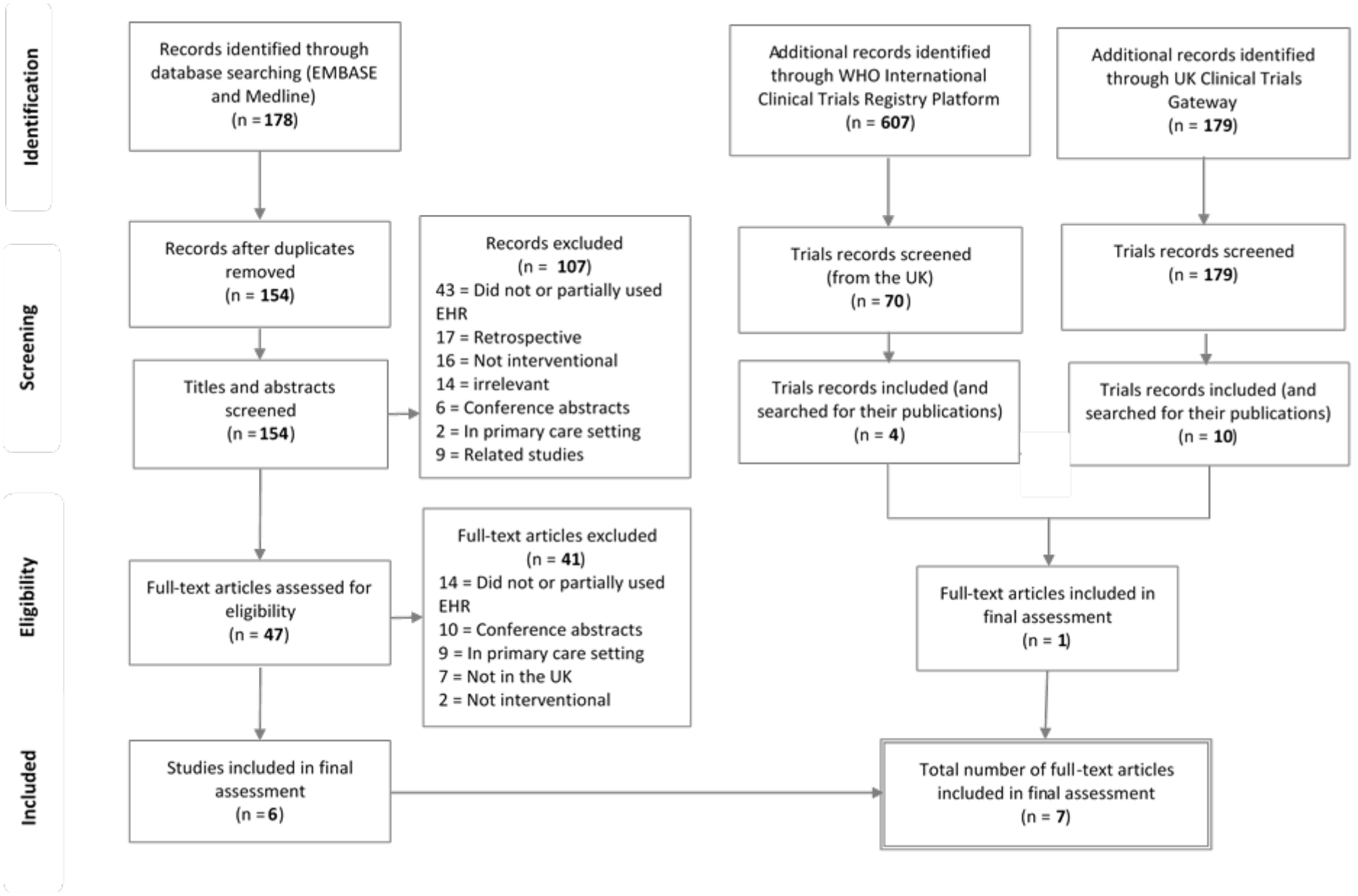
Williams	Secondary and follow-up outcomes	The Scottish national morbidity record with linkage to the General Registers Office, the Information Services Division of NHS Scotland and electronic patient health records	Not mentioned	Not mentioned	Manual	Participants were recruited from dedicated cardiology chest pain clinics where they were referred to with suspected angina with coronary heart disease
Wright-Hughes	Secondary and long-term outcomes	Accident and Emergency Departments (A&E), in-patient Hospital Episode Statistic (HES), Acute Trust records	Hospital attendance data obtained from A&E and inpatient HES datasets from NHS Digital. These data were augmented by directed hospital record searches, undertaken by masked researchers as required throughout the trial.	By the researchers	Manual	Young people were screened by a clinician at CAHMS after index self-harm episode for eligibility
Kravariti	None reported	Electronic clinical records system (electronic patient journey system Epjs)	Data extraction conducted using the Clinical Records Interactive Search (CRIS) system. CRIS provides authorised researchers with regulated access to over 250 000 fully anonymised ePJS records.	Research and clinical teams	Not mentioned	Not mentioned

Table 4: Reported experience of data extraction and collation by study investigators

Author	Comments on the processes of obtaining the data	Any comment in discussion on data collection/coding
Hewett	The trial team cleaned the data, removed duplication and resolved queries by discussion.	Data on patients admitted to the hospitals was collected by the hospital data system and masked to allocation.
Noble	For the trial phase, the CTU will provide regular reports on data quality to ensure the integrity of randomisation, to onitor the level of missing data and the timeliness of data entry and to check for llogical or inconsistent data. Data collection procedures will be monitored and source data verification against the paper data collection forms undertaken at regular intervals.	While the HES system is increasing in sophistication, it does not currently have a code to indicate visits related pecifically to epilepsy. However, in order to increase specificity and provide a more reasonable estimate of such visits, we shall utilise a broader code that does exist within the system so as to identify only visits related to a central nervous system condition (excluding stroke).
O'Connor	This nationally linked database allowed us to identify whether a patient was re-admitted to hospital anywhere in Scotland with self-harm at any time since their index episode. As NHS ISD is not yet able to routinely and reliably link emergency department admissions, we had to use the medical notes for all participants (using the TRAKcare system, which covers NHS Lothian) to identify whether any participant presented to the emergency department (and was subsequently discharged) with selfharm within 6 months of their index episode.	NHS ISD successfully linked 512 (99%) of 518 randomised participants from both groups (five in the intervention group vs one in the control group were not linked). We were able to identify emergency department re-presentations (via medical notes) for all patients.
Relton	none reported	All area-level data were collected routinely (and independently of the trial)

Williams	The Scottish national morbidity record with linkage to the General Registers Office was used to obtain information on the long-term outcomes of the study. This information also was obtained from the Information Services Division of NHS Scotland and, where appropriate, confirmed with review of the electronic patient health records.	none reported
Wright-Hughes	The primary outcome measure was obtained from Accident and Emergency Departments (A&E) and in-patient Hospital Episode Statistic (HES) data downloads from the NHS Health and Social Care Information Centre (HSCIC), which holds hospital attendance and admission data for hospitals across England.	Researchers searched Acute Trust records for episode details that were unclear from the central HES data, or for any hospital attendances for those participants who had not consented to the trial team providing their details to the HSCIC.
Kravariti	Ethical approval for CRIS as an anonymisation portal for secondary analyses was provided by an NHS research ethics committee in accordance with the Declaration of Helsinki, as well as by the Institute of Psychiatry's nstitutional Review Board.	To enable data extraction, the research team tagged all participant appointments in ePJS at the point of appointment entry using unmodifiable study- and arm-specific tags. These were later used as search terms in CRIS to extract the data.

Figure 1: PRISMA Flow Diagram



Acknowledgements

Sheena Visram - Quality Improvement Lead, ICHNT

Ben Glampson - Research Informatics Program Manager, ICHNT

Dr Ravi Aggarwal, Clinical Research Fellow, Imperial College London

Dr Bob Klaber - Deputy Medical Director, ICHNT

Prof Ara Darzi - Professor, Imperial College London

References

S A Julious, M J Horspool, S Davis, et al. PLEASANT: Preventing and Lessening Exacerbations of Asthma in School-age children Associated with a New Term – a cluster randomised controlled trial and economic evaluation. Health Technology Assessment, Dec 2016. No. 20.93.

van Staa T-P, Goldacre B, Gulliford M, et al. Pragmatic randomised trials using routine electronic health records: putting them to the test. BMJ. 2012;344:e55

van Staa T, Dyson L, McCann G, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. Health Technol Assess (Winchester, England). 2014;18(43):1 –146

Williams J, Cheung W, Cohen D, Hutchings H, Longo M. Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment. *Health Technol Assess* 2003;7(26)

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses:* The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Conflict of interest declaration

HR : no conflicts of interest

GM: no conflicts of interest

VC has received funding as an external consultant for DeepMind Technologies