USING ELECTRONIC HEALTH RECORD OR DISEASE REGISTRY DATA FOR CLINICAL TRIALS – A FRAMEWORK OF PRACTICE

(PROJECT No. 129740)

Dr. Stephen McCall, Dr. Chris Gale,
Associate Professor Edmund Juszczak

Ed.juszczak@npeu.ox.ac.uk

Foreword

Our original application for funding in 2017, as the title illustrates, was to investigate the use of electronic health record (EHR) or disease registry data for clinical trials.

We proposed to develop guidance for stakeholders – trialists, Clinical Trial Units, reviewers, funders, healthcare professionals, participants – outlining the additional components to be considered when designing, analysing and reporting a clinical trial using an EHR database or existing registry.

Our plan was outlined following the guidance for developers of health research reporting guidelines laid out by Moher et al., PLoS Medicine 2010 (http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000217); including:

- A systematic review of the literature and existing guidance regarding electronic health record (EHR) and registry trials (including seeking evidence on the quality or reporting and potential sources of bias in such studies) – work stream 1
- An online Delphi consensus process (to identify minimum reporting components for trials using EHR databases or disease registries) followed by a face-to-face consensus meeting with relevant stakeholders (to identify design, analysis and reporting components relevant to trials using EHR databases or disease registries) – work stream 2
- Development of a CONSORT extension for trials using EHR databases or disease registries – work stream 3

At the time we submitted our proposal, we became aware of a CONSORT Extension for Cohort-and Registry-embedded Trials that was registered on the EQUATOR network (http://www.equator-network.org/library/reporting-guidelines-under-development/#77). Our proposed extension was fundamentally different from this registered extension but complementary in that it was primarily concerned with the use of EHR data for clinical trials, rather than data collected through an established cohort, administrative dataset or disease registry.

Having identified this other project, we contacted the authors and rapidly became part of an international collaboration addressing the use of cohorts and routinely-collected data in clinical trials (including cohorts, administrative databases, disease registries and electronic health records), EJ being designated the lead on the electronic health records theme.

We have endeavoured to report on the one element of this collaboration (the largest), namely clinical trials using electronic health records, and apologise for any subsequent confusion, since it has been hard to divorce our theme from the overall project.

The final output, a work in progress, is a CONSORT statement extension for randomised controlled trials (RCTs) using cohorts and routinely-collected health data, which, of course, includes RCTs using electronic health records.

Contents

List of Tables

List of Figures

Abstract

Background: Randomised controlled trials (RCTs) are increasingly being conducted using existing sources of data such as cohorts, electronic health records, administrative databases and disease registries. Transparent and complete reporting of RCTs conducted using existing data sources requires inclusion of additional information. This reporting guideline is an extension of the 2010 version of the Consolidated Standards of Reporting Trials (CONSORT) Statement for RCTs using cohorts and routinely-collected health data.

Methods: A 'long list' of potential reporting items was identified through two methods: firstly, the additional items were identified from existing reporting guidelines, including the Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) and REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statements. Secondly, a scoping review of RCTs conducted in the last decade using cohorts and routinely-collected health data was undertaken and potential reporting items extracted. The 'long list' was entered into a three-stage Delphi consensus exercise of trialists and methodologists to assess the importance of each item for inclusion in the final CONSORT extension checklist, which was finalised at a face-to-face meeting of experts.

Results: A long list of 27 items was created and 125 experts registered for the three-round Delphi exercise (92, 77 and 62 experts participated in each round respectively). Consensus was reached on 21 out of 27 items in the Delphi exercise, and during the consensus meeting 8 additional items and 7 modified items were included in the final checklist. The checklist was disseminated and discussed through an invited session in May 2019 at the Society for Clinical Trials conference in New Orleans. Corresponding explanations and examples for each modified and additional item are being presently developed for publication and further dissemination.

Conclusion: We have produced a reporting guideline to facilitate transparent reporting of RCTs using cohorts and routinely-collected health data. Use of this guideline will assist evaluations of rigour and reproducibility, enhance understanding of the methodology, and make the results more useful for trialists, clinicians, journal editors, reviewers, guideline authors and funders.

Introduction

Large definitive randomised controlled trials (RCTs) are challenging to undertake; key challenges include recruitment (1) and extensive monitoring and regulatory requirements (2), both of which increase the resources required to complete RCTs. In response, new approaches to conducting RCTs have been developed including using existing data structures, such as cohorts and routinely-collected data. Routinely-collected data consists of three overlapping data systems: electronic health records (EHR), administrative databases and disease registries. These data structures offer the opportunity to conduct RCTs efficiently through automatic systems to identify potential participants, assess their eligibility, record consent, randomise, and collect all trial data (3).

The Consolidated Standards of Reporting Trials (CONSORT) 2010 Statement checklist is a 25-item instrument that was established to facilitate transparent and complete reporting of RCTs (4). A systematic review showed that RCTs published within journals that endorsed the CONSORT guideline had more complete reporting than journals that did not (5). The original statement was designed for parallel-group trials, and CONSORT extensions have been adapted to meet the reporting requirements of other RCT designs (6-8). RCTs conducted using cohorts and routinely-collected data have specific reporting requirements, including issues regarding data quality and the enrolment and consent process. As a result, a tailored reporting guideline is required to facilitate clear and transparent reporting of RCTs conducted using these data sources. The aim of this project is to develop and publish a CONSORT reporting extension for RCTs conducted using cohorts and routinely-collected data (9).

Methods

The overall process, timeline and the strategy for the development of a CONSORT extension for cohorts and routinely-collected data is illustrated in Figure 1 (9).

Project steering committee

This included experts in conducting RCTs using cohorts and routinely-collected data and RCT methodology. The core team, which undertook the work, included members from the University of Oxford (EJ, SM), Imperial College London (CG), McGill University (BT, MI and LK), Queen Mary's University London (CR), University of Basel (LH and KM), University of Western Ontario (MZ) and Örebro University (OF).

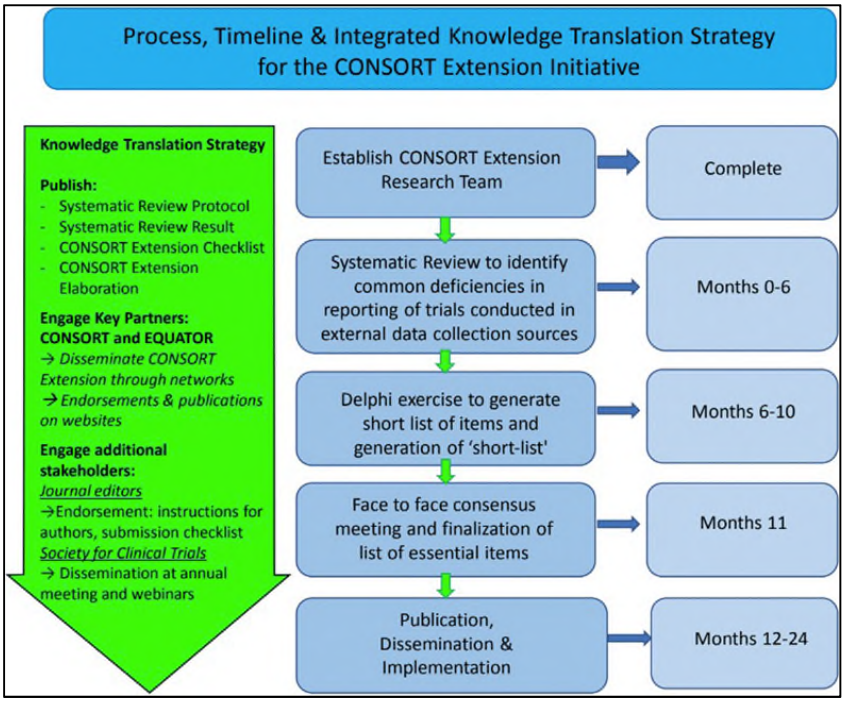
'Long list' creation

A long list of items was created using two methods. First, the original CONSORT checklist was used as a template for adapting existing items and identifying areas for additional items, which were extracted from guidelines such as the STROBE (10) and RECORD statements (11). Secondly, a scoping review was undertaken to identify additional reporting items from RCTs that use cohorts or existing data sources.

Scoping review

This was prospectively registered (12) and the objective was to identify reporting needs for trials conducted using cohorts and routinely-collected data (including EHRs, administrative databases and disease registries) and to find examples of good reporting. Reviews focusing on the four separate components were conducted by different teams; this report focuses on the EHR review which was led and undertaken by the team at NPEU (EJ, SM), University of Oxford and Imperial College London (CG). The review included protocols or reports of RCTs that had used the routinely-collected data for **both** identification/screening for participants and ascertainment of trial outcomes. Trial methodology papers that were relevant to these types of trials were also included in the review.

Figure 1. The process and timeline of the CONSORT extension development strategy.



Published in: Kwakkenbos, Linda, et al. 2018. Research integrity and peer review 3. P5. Reproduced under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/)

Search strategy

Separate searches were performed to identify publications describing methodology, trial protocols and results from RCTs that were conducted using (1) EHRs (2) registries (3) administrative databases or (4) cohorts. Searches were undertaken in Ovid MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE and EBM Reviews-Cochrane Methodology Registry (Final issue, third Quarter 2012). Searches were conducted covering 2007–18 which allowed the identification of recent publications. The references were imported from the database into Refworks and duplicates were removed. The references were then imported into the systematic review software DistillerSR (Evidence Partners, Ottawa, Canada, (13)). A coding manual, tailored to each data source to identify and code publications for inclusion, was developed and followed by independent coders.

The titles and abstracts were screened independently by two reviewers (MI, SJ). A liberal accelerated method, where titles and abstracts are screened by one reviewer and excluded articles are screened by a second reviewer, was used to identify articles for inclusion for full

text review. Reviewers were blinded to whether the other reviewer had already made a decision on any given title and abstract. Full texts were screened independently by two reviewers (SM and MI), and discordant cases were resolved by a third reviewer (BT or LK). At the full text screening, each reviewer indicated how the data source was used: not at all, data source used for identification of participants/recruitment, ascertainment of outcomes, both identification of participants/recruitment and outcomes, or for delivering the intervention. Publications reporting trials where an existing data source was used as part of delivering the intervention were identified in this review but were not included in the wider CONSORT scoping review, unless they also used an existing data source for both identification/screening for participants and ascertainment of trial outcomes.

Data extraction

Data were extracted only from RCTs that used the data source for both the recruitment of participants and ascertainment of outcomes. Using articles from the four separate reviews, potential gaps in reporting or relevant areas to the reporting of these RCTs provided the evidence for the creation of additional checklist items. In addition, these reviews identified modifications to existing 'long list' checklist items. Finally, publications were screened for potential examples of good reporting for each item on the 'long list'. Two investigators (SM and EJ) double-checked the 'long list' for redundancy (i.e. to check whether the suggested items were specific to cohorts and routinely-collected data – it was not our role or intention to modify or rewrite original CONSORT items pertinent to all trial designs) and duplication.

Additional scoping review

The scoping review was expanded to explore how EHRs were used within RCTs and to describe the characteristics of these RCTs. Using the results of the full text screening, trials that had either used the data source for identification of participants/recruitment, ascertainment of outcomes or for delivering the intervention were included. In a separate data extraction form, further information about the trials such as type of RCT, setting, location, sample size, intervention, comparator and outcome was collected. These results will be presented separately from the CONSORT scoping review.

Delphi exercise

Stakeholders

Key stakeholders were identified to participate in the three-round Delphi exercise (see Appendix A for details). A global invitation to trials and methodology experts was sent out using networks including the UKCRC CTU network and MRC trials methodology framework.

Three-round Delphi exercise

The 'long list' formed the basis for a three-round online Delphi consensus exercise to identify reporting items deemed essential for inclusion in the CONSORT extension. The Delphi participants were asked to score each proposed 'long list' item. Scoring comprised a 5-point Likert scale with a score of 1 or 2 deemed non-essential, 3 for further discussion and 4 or 5 deemed highly essential. Consensus was reached for an item when at least two-thirds of the Delphi participants rated it either essential or non-essential. If an item reached consensus, it was not entered into the next round of the Delphi exercise. Comments were invited from respondents to address whether they thought any other important reporting items had been omitted.

Face-to-face consensus meeting

A consensus meeting was held at Imperial College, London on 13–14 May 2019 (see Appendix A for list of stakeholders). All items were taken forward to the consensus meeting. Items that reached consensus were then ratified, modified or excluded at the meeting, and those items that had not reached consensus were reviewed for inclusion. In addition, items that were not included in the Delphi exercise were discussed and were either included, modified or excluded through consensus. At the consensus meeting, items that did not meet consensus went forward to a vote. For an item to be included in the final checklist, at least 80% of attendees had to be in favour of including the item.

Roles of collaborators

This project formed part of an international collaboration and each team focused on one particular routine data source. The team at the NPEU, University of Oxford and Imperial College London completed the systematic review on RCTs using EHRs and this review had the largest number of articles (~200 full texts) and will publish the results of the systematic review separately, alongside a baseline assessment of the quality of reporting assessed

against the new proposed CONSORT extension. The McGill team undertook the review on RCTs using cohort studies and conducted the whole administrative process for undertaking the Delphi exercise process and face-to-face meeting (the latter with help from the UK team). Collaborators from the University of Basel, University of Western Ontario and Örebro University focused on the administrative databases and registries. Steering committee members attended bi-weekly teleconferences to discuss progress of the scoping review, 'long list' items, Delphi results, administrative processes and logistics for the face-to-face consensus meeting. They also assisted in the identification of stakeholders and participated in the face-to-face consensus meeting. The NPEU, University of Oxford team and Imperial College London jointly led the development of the presentation at the Society for Clinical Trials, New Orleans with the Canadian team (BT, MI) and led the submission of an abstract to the International Clinical Trials Methodology Conference in Brighton in October 2019, which has been accepted for oral presentation.

Results and Conclusion

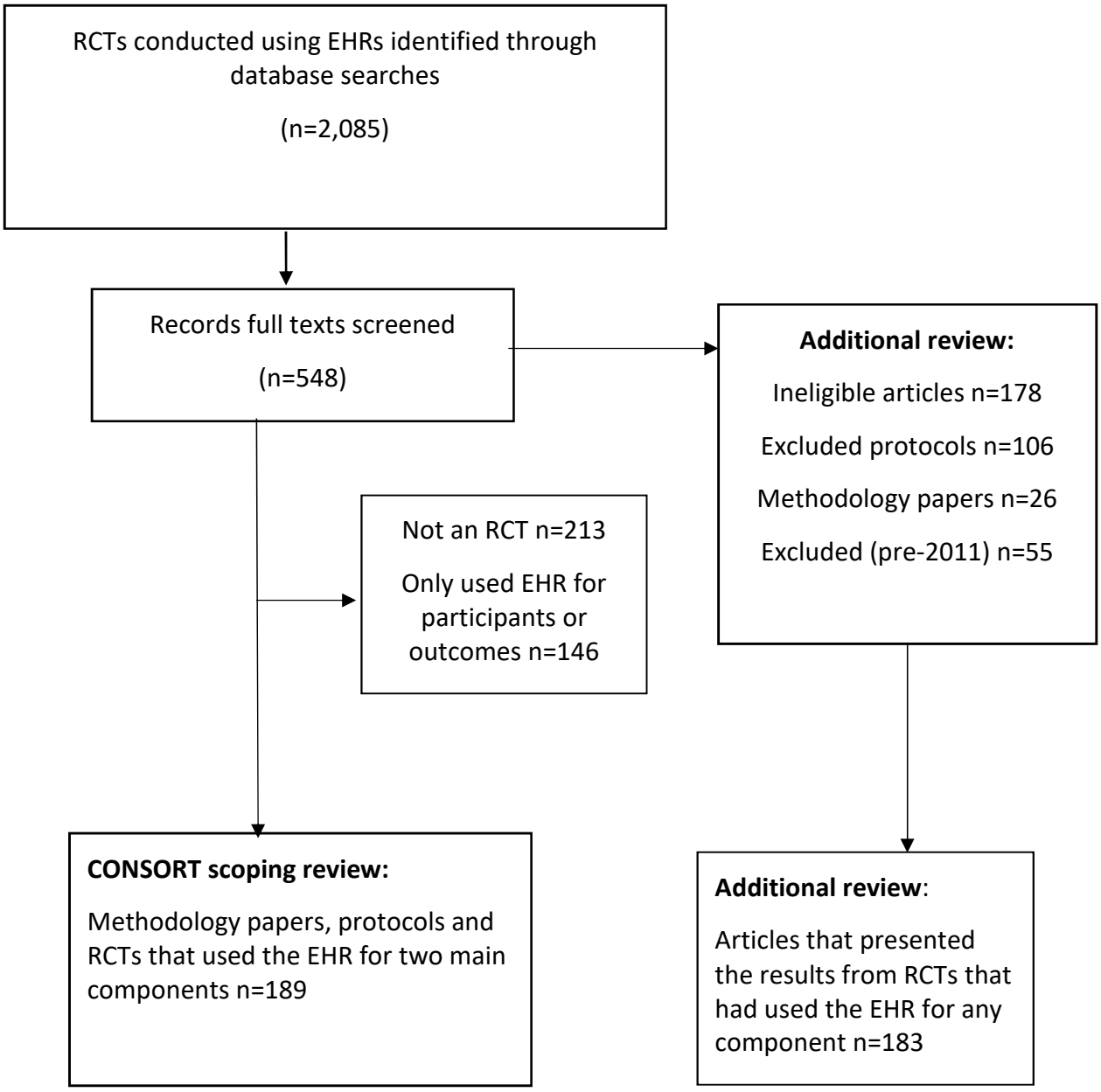
Scoping review for CONSORT extension

The results of the scoping review are presented in Table 1. The largest number of identified articles related to RCTs conducted using electronic health records. The derivation and screening of articles included in the review conducted by the NPEU, University of Oxford and Imperial College London team are presented in Figure 2. The 'long list' consisted of 7 modified and 20 additional items, which were derived from the existing CONSORT 2010, RECORD and STROBE guidelines or scoping review and these are presented in Table 3.

Table 1. Results of the scoping review for each data source

	Number of protocols or RCT results publications	Number of methodology papers
Review of RCTs conducted using the following data sources		
1. Electronic health records (Oxford-led)	169	20
2. Registries	12	11
3. Administrative databases	24	0
4. Cohorts	65	17

Figure 2. PRISMA flow chart of studies included in the review examining RCTs conducted using electronic health records.



Additional review

The breakdown of studies included in the wider additional review by 'use of the EHR' is presented in Figure 3. The Figure illustrates that nearly half of published EHR RCTs used the EHR for all three components of the trial (identification, intervention and outcomes). Further characteristics about these trials is presented in Table 2.

Figure 3. Overall characteristics of published articles identified and according to how the EHR was used in the trial

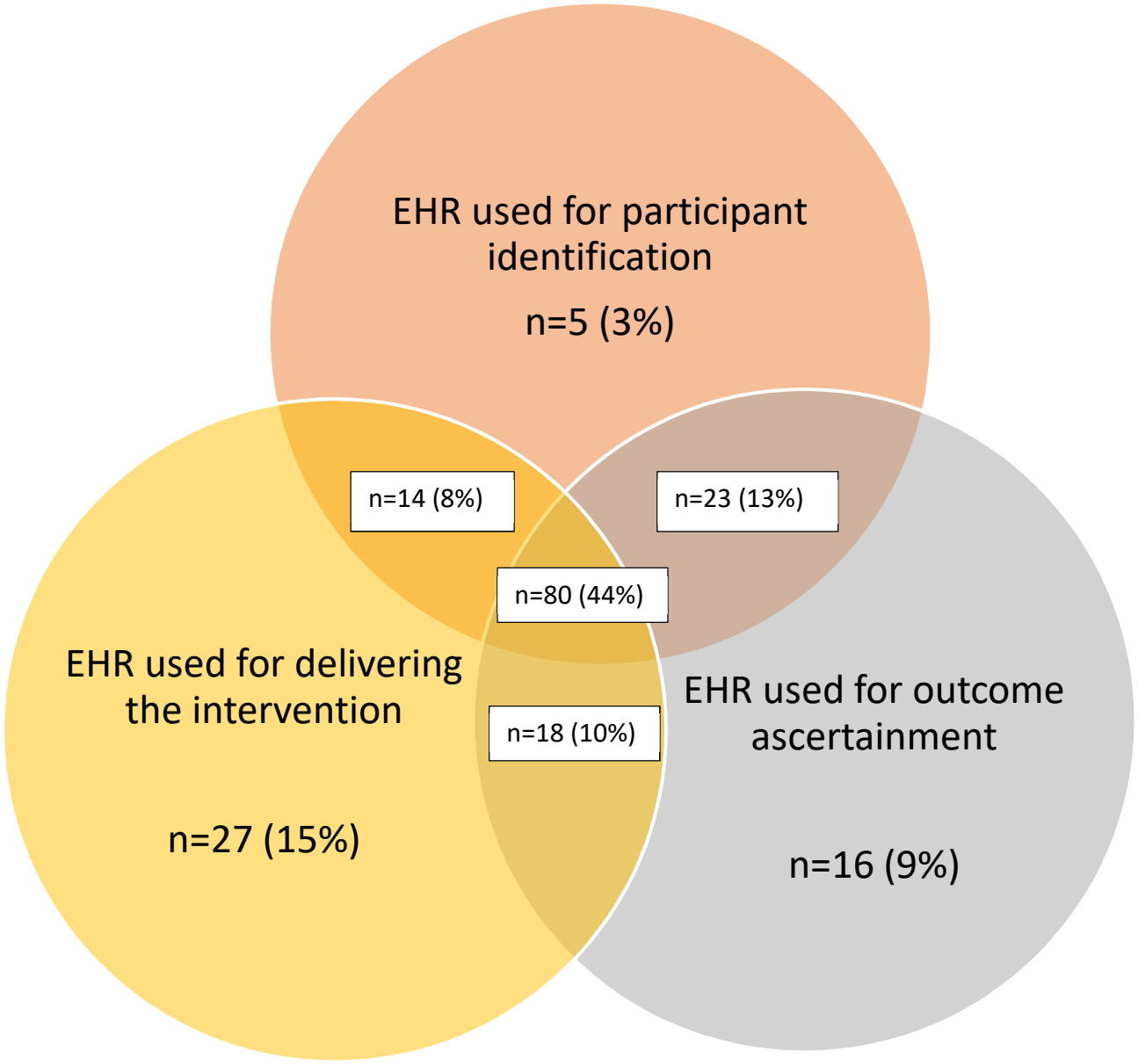


Table 2. Characteristics of trials conducted using electronic health records

	Number (%) of cluster randomised trials (n=84)		Number (%) of individually randomised trials (n=99)		Total (n=183)	
Setting						
Primary Care/Community medicine/Outpatient	67	(79.8)	67	(67.7)	134	(73.2)
Inpatient	9	(10.7)	16	(16.2)	25	(13.7)
Accident and Emergency	4	(4.8)	10	(10.1)	14	(7.7)
Pharmacy	3	(3.6)	5	(5.1)	8	(4.4)
Other	1	(1.2)	1	(1)	2	(1.1)
Country						
United Kingdom and Ireland	8	(9.5)	3	(3)	11	(6)
North America	65	(77.4)	78	(78.8)	143	(78.1)
Continental Europe	10	(11.9)	9	(9.1)	19	(10.4)
Rest of the world	1	(1.2)	9	(9.1)	10	(5.5)
Disease of interest						
CVD	15	(17.9)	17	(17.2)	32	(17.5)
Drug prescribing/use	8	(9.5)	2	(2)	10	(5.5)
Cancer	1	(1.2)	7	(7.1)	8	(4.4)
CKD	3	(3.6)	1	(1)	4	(2.2)
Mental Health	2	(2.4)	7	(7.1)	9	(4.9)
Diabetes	7	(8.3)	10	(10.1)	17	(9.3)
Respiratory/Gastro/Old age med	7	(8.3)	3	(3)	10	(5.5)
General medicine/health	2	(2.4)	6	(6.1)	8	(4.4)
Vaccinations	9	(10.7)	6	(6.1)	15	(8.2)
Infections and infectious diseases	2	(2.4)	3	(3)	5	(2.7)
Paediatrics	4	(4.8)	8	(8.1)	12	(6.6)
Risk factors	10	(11.9)	8	(8.1)	18	(9.8)
Other	14	(16.7)	21	(21.2)	35	(19.1)
Intervention						
Drug	1	(1.2)	2	(2)	3	(1.6)
Guideline/reminder-based	43	(51.2)	44	(44.4)	87	(47.5)
Multiple	6	(7.1)	1	(1)	7	(3.8)
Other	22	(26.2)	41	(41.4)	63	(34.4)
Screening	12	(14.3)	9	(9.1)	21	(11.5)
Surgical	0	(0)	2	(2)	2	(1.1)
Comparator						
Active comparison group	12	(14.3)	20	(20.2)	32	(17.5)
Usual care	71	(84.5)	74	(74.7)	145	(79.2)
Placebo	0	(0)	1	(1)	1	(0.5)
Unclear	1	(1.2)	4	(4)	5	(2.7)
Outcome						
Composite	2	(2.4)	3	(3)	5	(2.7)
Disease occurrence	2	(2.4)	1	(1)	3	(1.6)
Mortality	0	(0)	1	(1)	1	(0.5)
No primary outcome	4	(4.8)	12	(12.1)	16	(8.7)
Other	19	(22.6)	14	(14.1)	33	(18)
Self-reported	7	(8.3)	22	(22.2)	29	(15.8)
Surrogate	14	(16.7)	17	(17.2)	31	(16.9)
Uptake of treatment or service e.g. drug or vaccine	36	(42.9)	29	(29.3)	65	(35.5)

EHR for intervention*						
EHR not used for intervention	19	(22.6)	25	(25.3)	44	(24)
Clinical decision support	57	(67.9)	30	(30.3)	87	(47.5)
Personal health record	3	(3.6)	20	(20.2)	23	(12.6)
Telehealth	2	(2.4)	17	(17.2)	19	(10.4)
Other	3	(3.6)	7	(7.1)	10	(5.5)
EHR for used for primary outcome						
No	15	(17.9)	38	(38.4)	53	(29)
Not clear	5	(6)	7	(7.1)	12	(6.6)
Yes	64	(76.2)	54	(54.5)	118	(64.5)
Sample size						
Number of clusters (median and IQR)	28	(15-59)				
Total number of clusters	6,020					
Number of participants (median and IQR)	4,447	(613-20,904)	400	(123-1,999)		
Total number of participants	2,312,144		275,603			

IQR denotes interquartile range (25th percentile to 75th percentile)

**Definitions adopted from Hemkens and Mc Cord, 2019, CMAJ (3)*

Clinical decision support tools use software that mines information from the EHR and triggers a warning or message to the clinician when a code, disease or phrase is entered. It will often contain a warning message, e.g. in a trial 'antibiotic overuse' or based on a recommendation from a guideline.

Telehealth are telecommunication systems that can communicate between the EHR and the patient.

Personal health records are systems that allow patients to access their EHR. The patient can add and receive information to this system and access their test results. The transfer of information occurs in both directions and is directly connected to the EHR.

Other interventions that use the EHR include: (i) an integrated system that aids with prescriptions to prevent mistakes, drug interactions and the correct drug usage, (ii) electronic patient reported outcomes (ePROs). This is where patients can electronically report their outcomes through a mobile application or website; however, the patient does not have access to the EHR itself. If this information is uploaded onto an EHR then it would be considered as an EHR based intervention.

There were 84 (45%) cluster trials and 99 individually randomised trials (54%). The vast majority were conducted in North America (78%), set in primary care, community medicine or outpatient clinics (73%) and the most common disease of interest was cardiovascular disease (~18%). Around half of the interventions were guideline or reminder-based systems and 4 out of 5 trials used 'usual care' as the comparator. If the EHR was used for delivering the intervention, a clinical decision support tool was the most popular method, however this differed depending on the whether it was a cluster or individually randomised trial. The most common outcome was the uptake of a treatment or service (~1 in 3) and nearly 2 in 3 trials used the EHR to identify the primary outcome. In terms of sample size, for cluster trials, the average number of participants was 4,447 (interquartile range [IQR] 613 to 20,904) with 28 clusters (IQR 15 to 59); for individually randomised trials, the average number of participants was much smaller, at 400 (IQR 123 to 1,999).

Delphi exercise

The 'long list' of 27 items was assessed by 125 experts in the three-round Delphi exercise, where 92, 77 and 62 experts participated in each round, respectively. In stage 1, with a response rate of 72% from the invited experts, 14 out of 27 items (52%) reached consensus. In stage 2, the remaining 13 items were rated by respondents and 2 (15%) items reached consensus. In the final stage, consensus for inclusion was reached for a further 5 items. This gave a total of 21 out of 27 items reaching consensus in the Delphi process and no item reached consensus for exclusion. The consensus status for each item in 'long list' is presented in Table 3 (under embargo since being prepared for publication of a methods paper).

Face-to-face consensus meeting

A face-to-face consensus meeting was held at Imperial College London from 13–14 May 2019. The attendees comprised the Steering Committee and CONSORT extension group (see Acknowledgement section). During the face-to-face consensus meeting, out of the original 27 'long list' items, 8 additional items and 7 modified items were agreed to be included in the final checklist for the CONSORT extension on RCTs conducted using cohorts and routinely-collected data (Table 4 – checklist under embargo since being prepared for publication alongside an Explanation and Elaboration document).

Future value of this work

RCTs using routinely-collected data are increasing and this CONSORT extension is timely as it provides a benchmark of reporting for these type of RCTs, which will be normal practice in the near future. Thus this work funded by the NIHR will have future impact on the quality of reporting of these RCTs globally. In addition, transparency of reporting has the secondary impact of improving the quality of RCTs. This NIHR-funded project is a vital component in reducing research waste.

Table 3. The 'long list' of additional and modified items that were assessed in the Delphi exercise for whether they met consensus [Under embargo]

ORIGINAL CONSORT Item			Suggested modified or additional extension items	Consensus for inclusion status
Title and abstract				
	1a	Identification as a randomised trial in the title	Identification as a randomised trial in the title, including that it was a trial conducted using a cohort or routinely-collected source of data (Modified)	Not reached
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)		
			The source(s) of data used to conduct the trial should be specified in the abstract (Additional)	Reached
			If linkage between multiple sources of data was conducted for the study, this should be clearly stated in the abstract (Additional)	Not reached
			The proportion of participants offered and the proportion that accepted the intervention should be reported (for trials conducted using the cohort multiple RCT design) (Additional)	Reached
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale		
	2b	Specific objectives or hypotheses		
Methods				

Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Description of trial design (such as parallel, factorial) including allocation ratio, the source(s) of data used to conduct the trial (such as cohort, registry) and how the data are used within the trial (such as identification of eligible trial participants, trial outcomes) (Modified)	Reached
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		
Source(s) of data			Description of the source(s) of data used to conduct the trial, including the setting, locations, relevant dates, periods of recruitment, follow-up, and data collection (Additional)	Reached
			Describe indicators of the quality of the source(s) of data used to conduct the trial including what types of quality checks have been performed and the entity responsible for the data (Additional)	Reached
			Describe modifications to the data collected in the source(s) of data used to conduct the trial, such as adding data items, if applicable (Additional)	Reached
			Describe additional sources of data used to conduct the trial, if any (Additional)	Reached
			Give the eligibility criteria, the sources and methods of selection of participants, and methods of follow-up (for trials conducted using cohorts or registries) (Additional)	Reached
			Detail any use of record linkage across sources of data, the methods of linkage and methods of quality evaluation, if applicable (Additional)	Reached
			Describe if (and how) participants were informed about the potential use of their data in randomised trials (Additional)	Not reached

Trial participants	4a	Eligibility criteria for participants	Eligibility criteria for trial participants (Modified)	Reached
			Provide details of how eligible clusters/participants were identified from the source(s) of data used to conduct the trial (Additional)	Reached
	4b	Settings and locations where the data were collected	Settings and locations where the trial data were collected (Modified)	Reached
			Details of information provided to participants from the source(s) of data who are selected for recruitment or inclusion in the trial, including any differences in information provided across trial arms (Additional)	Not reached
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered		
			Describe how the source(s) of data was used to implement the intervention, if applicable (e.g., for trials conducted using electronic health records) (Additional)	Reached
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed		
			Provide source(s) of data for each outcome (Additional)	Reached
			Provide a list of codes and algorithms used to define (and/or derive) the outcomes as supplementary information, including validation, if applicable (Additional)	Not reached

			Detail any adjudication or external validation of data items from the source(s) of data used to conduct the trial, if applicable (Additional)	Reached
	6b	Any changes to trial outcomes after the trial commenced, with reasons		
Sample size	7a	How sample size was determined		
	7b	When applicable, explanation of any interim analyses and stopping guidelines		
Randomisation:				
Sequence generation	8a	Method used to generate the random allocation sequence		
	8b	Type of randomisation; details of any restriction (such as blocking and block size)		
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Mechanism used to implement the random allocation sequence, describing any steps taken to conceal the sequence until interventions were assigned, such as using automated random sequence generation concealed within source(s) of data (Modified)	Reached
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions		
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how		

	11b	If relevant, description of the similarity of interventions		
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes		
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		
Results				
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	Describe in detail the numbers of clusters/participants in the source(s) of data used to conduct the trial, number screened for eligibility, randomly assigned, offered and accepted interventions (e.g., cohort multiple RCTs), received intended treatment, and analysed for the primary outcome (Modified)	Reached
			Describe any linkage of multiple sources of data, including the number of clusters/participants successfully linked (Additional)	Reached
	13b	For each group, losses and exclusions after randomisation, together with reasons		
Recruitment	14a	Dates defining the periods of recruitment and follow-up		
	14b	Why the trial ended or was stopped		
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group		

			A table showing baseline demographic and clinical characteristics for eligible participants who participated in the trial and those who did not (Additional)	Reached
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups		
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)		
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory		
			If outcomes for eligible patients in the existing source(s) of data who were not included in the trial are known, they should be reported (Additional)	Not reached
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)		
Discussion				

Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		
			Discuss the implications of using data that were not created or collected to answer the specific research question(s) (Additional)	Reached
Generalisability	21	Generalisability (external validity, applicability) of the trial findings		
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		
Other information				
Registration	23	Registration number and name of trial registry		
Protocol	24	Where the full trial protocol can be accessed, if available		
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	Sources of funding and other support for the trial and the existing source(s) of data, role of funders (Modified)	Reached

Table 4. The checklist of the CONSORT Extension for trials conducted using cohorts and routinely-collected health data [Under embargo]

ORIGINAL CONSORT Item			Suggested modified or additional extension items
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
	1c		The cohort or routinely-collected database(s) used to conduct the trial should be specified in the abstract (Additional)
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Description of trial design (such as parallel, factorial) including allocation ratio, the cohort or routinely-collected database(s) used to conduct the trial (such as cohort, registry) and how the data were used within the trial (such as identification of eligible trial participants, trial outcomes) (Modified)
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Cohort or Routinely-collected database	4a		Name and description of the cohort or routinely-collected database(s) used to conduct the trial, including information on the setting (such as primary care), locations, and dates, (such as periods of recruitment, follow-up, and data collection) (Additional)
	4b		Eligibility criteria for participants in the cohort or routinely-collected database(s) (Additional)

	4c		State whether the study included person-level, institutional-level, or other data linkage across two or more databases and, if so, linkage techniques and methods used to evaluate completeness and accuracy of linkage (Additional)
Trial participants	5a	Eligibility criteria for participants	Eligibility criteria for trial participants, including information on how to access the list of codes and algorithms used to identify eligible participants, information on accuracy and completeness of data used to ascertain eligibility, and methods used to validate accuracy and completeness (e.g., monitoring, adjudication), if applicable (Modified)
	5b	Settings and locations where the data were collected	
	5c		Describe whether and how consent was obtained (Additional)
Interventions	6	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	7a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	Completely defined pre-specified primary and secondary outcome measures, including how and when they were ascertained and the cohort or routinely-collected database(s) used to ascertain each outcome (Modified)
	7b		Information on how to access the list of codes and algorithms used to define or derive the outcomes from the cohort or routinely-collected database(s) used to conduct the trial, information on accuracy and completeness of outcome variables, and methods used to validate accuracy and completeness (e.g., monitoring, adjudication), if applicable. (Additional)
	7c	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	8a	How sample size was determined	
	8b	When applicable, explanation of any interim analyses and stopping guidelines	

Randomisation:			
Sequence generation	9a	Method used to generate the random allocation sequence	
	9b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	10	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Mechanism used to implement the random allocation sequence (such as embedding the random allocation sequence within the cohort or routinely-collected database(s)), describing any steps taken to conceal the sequence until interventions were assigned (Modified)
Implementation	11	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	12a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	12b	If relevant, description of the similarity of interventions	
Statistical methods	13a	Statistical methods used to compare groups for primary and secondary outcomes	
	13b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	14a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the number of participants in the cohort or routinely-collected database(s) used to conduct the trial and the numbers screened for eligibility, randomly assigned, offered and accepted interventions (e.g., cohort multiple RCTs), received intended treatment, and analysed for the primary outcome (Modified)

	14b	For each group, losses and exclusions after randomisation, together with reasons
Recruitment	15a	Dates defining the periods of recruitment and follow-up
	15b	Why the trial ended or was stopped
Baseline data	16	A table showing baseline demographic and clinical characteristics for each group
Numbers analysed	17	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
Outcomes and estimation	18a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)
	18b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended
Ancillary analyses	19	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory
Harms	20	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)
Discussion		

Limitations	21	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	22	Generalisability (external validity, applicability) of the trial findings	
Interpretation	23	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence, including the implications of using data that were not collected to answer the specific research question (Modified)
Other information			
Registration	24	Registration number and name of trial registry	
Protocol	25	Where the full trial protocol can be accessed, if available	
Funding	26	Sources of funding and other support (such as supply of drugs), role of funders	Sources of funding and other support for both the trial and the cohort or routinely-collected database(s), role of funders (Modified)

Dissemination

Throughout this project, outputs have been published and disseminated and several publications are also planned (see Table 5). Protocols describing the CONSORT extension and scoping review have been published (9, 12). The systematic review examining RCTs that use EHRs will be published alongside a baseline assessment of the quality of reporting of these trials. The ultimate intended output of this project, the CONSORT extension checklist and corresponding explanation and elaboration document, will be published in an Open Access journal. It will also be made freely available through the EQUATOR network and CONSORT statement extension websites. Furthermore, the learning points about the methodology used for the CONSORT extension will also be published.

It is anticipated that an extension of the CONSORT statement would encourage improved conduct and reporting of these types of trials. As a by-product, it is likely that this body of work will encourage transparent reporting, which has the potential to reduce research waste.

Publications

Table 5. List of publications and planned outputs

Protocols
Kwakkenbos L, **Juszczak E**, Hemkens LG, Sampson M, Fröbert O, Relton C, **Gale C**, Zwarenstein M, Langan SM, Moher D, Boutron I, Ravaud P, Campbell MK, Mc Cord KA, van Staa TP, Thabane L, Uher R, Verkooijen HM, Benchimol EI, Erlinge D, Sauvé M, Torgerson D, Thombs BD. Protocol for the development of a CONSORT extension for RCTs using cohorts and routinely-collected health data. *Res Integr Peer Rev* 2018;3:9. Kwakkenbos L, Imran M, McCord KA, Sampson M, Fröbert O, **Gale C**, Hemkens LG, Langan SM, Moher D, Relton C, Zwarenstein M, Benchimol EI, Boutron I, Campbell MK, Erlinge D, Jawad S, Ravaud P, Rice DB, Sauve M, van Staa TP, Thabane L, Uher R, Verkooijen HM, **Juszczak E**, Thombs BD. Protocol for a scoping review to support development of a CONSORT extension for randomised controlled trials using cohorts and routinely-collected health data. *BMJ Open* 2018; 8:e025266.
Planned publications
Thombs, B., **Juszczak, E., Gale, C.,** Imran, M., Kwakkenbos, L., **McCall, S**., *et al*. CONSORT Extension for trials conducted using cohorts and routinely-collected health data. **McCall, S.,** Imran, M., Hemkens, L.G., Thombs, B., **Gale, C., Juszczak, E.,** Protocol to describe randomised controlled trials conducted using electronic health records with an assessment of completeness and transparency of reporting. **McCall, S.,** Imran, M., Hemkens, L.G., Thombs, B., **Gale, C., Juszczak, E.,** A descriptive study and assessment of completeness and transparency of reporting of randomised controlled trials conducted using electronic health records. Imran, M., **McCall, S.,** Hemkens, L.G., Thombs, B., **Gale, C., Juszczak, E.,** The results from a scoping review and Delphi process for the development of a CONSORT extension for trials conducted using cohorts and routinely-collected health data.

Conference presentations
Relton C., on behalf of CONSORT steering committee. Rethinking the architecture of pragmatic trials: A review of trials within both researcher generated and routine healthcare data structures. Society for Clinical Trials, Oregon, May 2018. **McCall S.,** on behalf of CONSORT steering committee. CONSORT extension for trials conducted using routinely-collected data. Nuffield Department of Population Health Annual Symposium, University of Oxford. March 2019. Campbell, M., **McCall, S., Gale, C., Juszczak, E.,** Thombs, B., Thabane, L., on behalf of CONSORT steering committee. CONSORT Extension for trials conducted using cohorts and routinely-collected health data. Society for Clinical Trials, New Orleans, May 2019. **Gale, C., Juszczak, E.,** on behalf of CONSORT steering committee. CONSORT Extension for trials conducted using cohorts and routinely-collected health data. International Clinical Trials Methodology Conference, October 2019.

Acknowledgements

Contribution of authors

Stephen McCall (NIHR-funded research associate; NPEU Clinical Trials Unit, Nuffield Department of Population Health, University of Oxford; systematic reviews and epidemiologist), Chris Gale (Co-investigator; Imperial College London; Clinical Trialist) and Edmund Juszczak (Principal Investigator; Director NPEU Clinical Trials Unit, Nuffield Department of Population Health, University of Oxford; Associate Professor of Clinical Trials Methodology and Statistics in Medicine). SM contributed to identifying stakeholders, was responsible for the completion of the EHR scoping review, identification of long list items and good examples of reporting, assisted in the administration for the Delphi exercise and contributed at the face-to-face meeting. EJ/CG secured the funding for this work and contributed to the grant application to the Canadian Institutes of Health Research; identified stakeholders; supervised the completion of the EHR scoping review; contributed to the identification of long list items and good examples of reporting; contributed to the real time interpretation of the Delphi exercise and contributed (EJ co-convened and presented, CG chaired a session and presented) at the face-to-face consensus meeting. EJ/CG/SMcCall presented the findings at the SCT conference session in New Orleans.

Steering committee

Ole Fröbert, Chris Gale, Lars Hemkens, Mahrukh Imran, Edmund Juszczak, Linda Kwakkenbos, Sinéad Langan, Stephen McCall, Clare Relton, Brett Thombs and Merrick Zwarenstein.

CONSORT extension group

Eric I Benchimol, Isabelle Boutron, Marion K Campbell, David Erlinge, Ole Fröbert, Chris Gale, Lars G Hemkens, Edmund Juszczak, Linda Kwakkenbos, Clare Relton, Sinéad M Langan, Kimberly A Mc Cord, David Moher, Philippe Ravaud, Margaret Sampson, Maureen Sauvé, Tjeerd P van Staa, Lehana Thabane, Brett D Thombs, David Torgerson, Rudolf Uher, Helena M Verkooijen and Merrick Zwarenstein.

Independent screening of titles and abstracts

We would also like to thank Sena Jawad (Imperial College London) for her contribution to the systematic review.

References

1. McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials. 2006;7(1):9.
2. Djurisic S, Rath A, Gaber S, Garattini S, Bertele V, Ngwabyt S-N, et al. Barriers to the conduct of randomised clinical trials within all disease areas. Trials. 2017;18(1):360.
3. Mc Cord KA, Ewald H, Ladanie A, Briel M, Speich B, Bucher HC, et al. Current use and costs of electronic health records for clinical trial research: a descriptive study. CMAJ open. 2019;7(1):E23.
4. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux P, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010;63(8):e1-e37.
5. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review a. Systematic reviews. 2012;1(1):60.
6. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot and feasibility studies. 2016;2(1):64.
7. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661.
8. Juszczak E, Altman DG, Hopewell S, K. S. Reporting of multi-arm parallel group randomized trials: extension of the CONSORT 2010 statement. JAMA. 2019;321(16):1610-20.
9. Kwakkenbos L, Juszczak E, Hemkens LG, Sampson M, Fröbert O, Relton C, et al. Protocol for the development of a CONSORT extension for RCTs using cohorts and routinely-collected health data. Research integrity and peer review. 2018;3(1):9.
10. Vandenbroucke J, Elm v, Altman D, Gøtzsche P, Mulrow C, Pocock S, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4:e297.
11. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885.
12. Kwakkenbos L, Imran M, McCord KA, Sampson M, Fröbert O, Gale C, et al. Protocol for a scoping review to support development of a CONSORT extension for randomised controlled trials using cohorts and routinely-collected health data. BMJ open. 2018;8(8):e025266.
13. Distiller S. Data Management Software. Ottawa, ON: Evidence Partners. 2015.

Appendix A

Stakeholders in the Delphi exercise

Trialists, participants/patients and public involvement, epidemiologists, healthcare professionals, journal editors, reviewers, CONSORT steering group, Clinical Trials Unit directors and trial methodologists/statisticians from the United Kingdom Clinical Research Collaboration, funders and newsletter circulations (e.g. to members of the Medical Research Council Clinical Trials Methodology Hub).

Participants in the face-to-face consensus meeting

Cohort, EHR, administrative database and registry RCT trialists, CONSORT steering committee, other trialists, PPI, librarian, journal editor, epidemiologists and routinely-collected data experts.

Conflict of interest declaration

EJ's institution, the University of Oxford, received funding from the National Institute of Health Research during the conduct of the study. EJ also declares membership of HTA General Board at that time. CG and SMcC have nothing to declare.

Word count: 15+844+1126+56+474+155 = 2,670 (sections 1-6 excluding Tables and Figures).

Total word count: 7,621