**Title: Appropriate statistical methods are infrequently used in cluster randomised crossover trials: Results from a systematic review**

**Authors:** Sarah J Arnup1\*, Andrew B Forbes1, Brennan C Kahan2, Katy E Morgan3, Joanne E McKenzie1

**Affiliations and emails:**

1. School of Public Health and Preventive Medicine, Monash University, The Alfred Centre, Melbourne, Victoria, 3004, Australia
2. Pragmatic Clinical Trials Unit, Queen Mary University of London, 58 Turner St, E1 2AB, London, UK
3. Medical Statistics Department, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

**Corresponding author:**

**\* Sarah J Arnup**

Email: sarah.arnup@monash.edu

Address: School of Public Health and Preventive Medicine, Monash University, The Alfred Centre, Melbourne, Victoria, 3004, Australia

Phone: +61 3 9903 0211

**Abstract**

***Objective***

To assess the design and statistical methods used in cluster randomised crossover trials (CRXO).

***Study design and setting***

We undertook a systematic review of CRXO trials. Searches of MEDLINE, EMBASE, and CINAHL Plus; and citation searches of CRXO methodological articles were conducted to December 2014. We extracted data on design characteristics, and statistical methods for sample size, data analysis, and handling of missing data.

***Results***

Ninety-one trials including 139 endpoint analyses met the inclusion criteria. Trials had a median of nine clusters (IQR 4-21); median cluster-period size of 30 individuals (IQR 14-77); 58 (69%) trials had two periods; and 27 trials (30%) included the same individuals in all periods. A rationale for the design was reported in only 25 trials (27%). A sample size justification was provided in 53 (58%) trials. Only nine (10%) trials accounted appropriately for the design in their sample size calculation. Ten of the 12 cluster-level analyses used a method that accounted for the clustering and multiple-period aspects of the design. In contrast, only 4 of the 127 individual-level analyses used a potentially appropriate method.

***Conclusions***

There is a need for improved application of appropriate analysis and sample size methods, and reporting, in CRXO trials.

**Keywords**: Cluster randomised crossover trial, crossover, cluster, sample size, design, statistical analysis

**What is new?**

**Key findings**

* Reporting of the rationale for using CRXO trials was uncommon, despite this being a recommended reporting item for cluster randomised trials[1]. Sample size calculations were commonly not reported, and only a minority of CRXO trials used sample size methods that appropriately accounted for the design.
* Only rarely did the employed statistical methods account for the design; that is, adjust for the clustering and multiple period aspects.

**What this adds to what is known**

* This is the first systematic review of CRXO trials. The results of this review provide a comprehensive assessment of the design characteristics, statistical methods for sample size, data analysis, and handling of missing data in CRXO trials.

**What is the implication, what should be changed**

* Trialists need to account for both the cluster randomisation and multiple period aspects of the design in sample size calculations and statistical analyses. Methods and assumptions need to be clearly reported and justified.
* The development of reporting guidelines for CRXO trials is needed to facilitate clearer and complete reporting.

**1. Introduction**

The cluster randomised crossover (CRXO) design is gaining popularity in settings where cluster randomisation is required, but the parallel group cluster randomised design is not feasible because the required number of clusters is prohibitively large[2, 3]. In the CRXO design, hospitals, schools or other groups of people (“clusters”) are randomly assigned to a sequence of interventions. Each cluster receives each intervention at least once in a separate period of time, leading to the formation of “cluster-periods” [4, 5]. Within each cluster, each cluster-period may contain a repeated-cross-section of different individuals, a cohort of the same individuals who are followed over time, or a mixture of the same and different individuals [6].

This design differs from the parallel group cluster randomized design and the individually randomized crossover design. In the parallel group cluster randomised design [7], each cluster is assigned only a single intervention, rather than a sequence of interventions. Each cluster therefore contains a single cross-section of different individuals. In the individually randomised crossover design [8], a cohort of individuals, rather than a series of clusters of individuals, are randomly assigned to a sequence of interventions. We refer the reader to Hooper & Bourke (2015) for examples of other cluster designs conducted over multiple periods [9].

In both the individually randomised crossover design and CRXO design, randomisation of the intervention sequence serves to control for period effects (i.e. changes that occur over time that are unrelated to the intervention); and a key requirement of the two period design is that the effect of an intervention given in one period does not carry over into the next period [5, 8]. In CRXO designs where cluster-periods contain different individuals, the potential for carryover is limited because any carryover can only take place at the cluster level. However in CRXO designs where the same individuals are followed over time, carryover can also take place at individual subject level, and therefore its potential is similar to that in individually randomized crossover designs.

The efficiency of a CRXO trial relative to an individually randomised trial or a parallel group cluster randomised trial depends on the relationship between the outcomes from individuals within and between each cluster-period [6]. Individuals within a cluster tend to have more similar outcomes than individuals across clusters[7]. For example, due to differences in case-mix between patients presenting to different hospitals, patients in the same hospital may have more similar outcomes than patients in other hospitals. Likewise, individuals within a cluster-period tend to have more similar outcomes than individuals in different clusters. This similarity is typically measured by the within-cluster within-period intracluster correlation (ICC) [3-6, 10]. This tendency for similar outcomes increases the uncertainty in the estimation of the effect of each intervention compared to outcomes that are independent.

If the environment of the cluster remains similar over time, then the outcomes of individuals within each cluster across different cluster-periods tend to be similar also. This tendency is typically measured by the within-cluster between-period ICC[3, 5, 6, 10]. By comparing the interventions within-cluster, the cluster specific variation is removed from the comparison, and the uncertainty of the difference between interventions is decreased when there is a positive within-cluster between-period ICC [5]. Therefore the crossover element of the design can offset the loss of precision arising from cluster randomisation.

In the analysis of data from a parallel group cluster randomised trial, it is recognised that the analysis must account for the correlation within clusters to correctly estimate the uncertainty in the intervention effect, for example, by including the cluster unit of randomisation as a random effect in a generalised linear model (GLM) (e.g. Eldridge 2012). However it is unclear whether trialists recognise that both the within-cluster within-period and the within-cluster between-period ICCs must be appropriately incorporated into sample size calculations and analyses to yield appropriate sample sizes and intervention effects with the correct standard errors in CRXO trials.

There have only been limited reviews examining the application and use of analytical methods for CRXO trials. These reviews have taken place in the introductory sections of methodological papers with the purpose of illustrating the design and highlighting the need for appropriate methods of analysis[4, 5, 10, 11]. Therefore we used systematic review methodology to examine the settings, design characteristics, justifications for using the design, quality of reporting, and sample size and analysis methods of trials that have used the CRXO design[12]. In this paper, we focus on the design characteristics; statistical methods for sample size and data analysis, and the appropriateness of those methods; and the completeness of reporting of the statistical methods.

We begin with a brief review of recommended sample size and analysis methods for CRXO trials in Section 2. In Section 3 we outline the systematic review methods. Results are presented in Section 4, and discussed in Section 5.

**2. Brief review of sample size and analysis methods for CRXO trials**

Only limited methodological research has been published to guide trialists in performing sample size calculations for CRXO trials. Giraudeau et al. and Donner et al. derived an inflation factor for a two-period two-group CRXO trial relative to an individually randomised trial, when the cluster-period sizes are assumed to be equal, and this was extended to incorporate period effects and unequal cluster-period sizes by Forbes et. al[3, 10, 13].

CRXO trials can be analysed at the level of the individual or at the level of the cluster. The statistical challenges for the analysis of CRXO trials differ according to whether an individual-level or cluster-level analysis method is chosen, however, the target parameter of the analysis remains the intervention effect at the individual-level [4, 14].

For individual-level analyses, the use of mixed effects models with continuous outcomes that include random or fixed effects for cluster and cluster-by-period effects have been shown to generally perform well in numerical simulations[4]. The performance of logistic mixed models for binary outcomes has only recently received attention with initial results indicating that poor performance may occur even with 50 or more clusters depending on the ICC values (Morgan et al., unpublished result). Other outcome types, including count and time-to-event data, have not been explored. Generalised estimating equations (GEE) are also used to analyse clustered data[15]. In CRXO trials, accounting for both ICCs requires the inclusion of robust standard errors (possibly with adjustments for small sample variance estimation[16]), or extensions which model the patterned correlation structure within clusters across multiple periods explicitly[3].

Turner et al. recommended a simple approach using a cluster-level analysis for continuous outcomes collected at the individual-level, however this method can also be applied to other outcome types[4]. In a cluster-level analysis, the available data on each individual from each cluster-period are aggregated into a single measure, and for each cluster, the relevant difference between interventions is constructed. Aggregating the data in each cluster-period accounts for the within-cluster within-period ICC, and comparing the cluster-period summaries within-cluster accounts for the within-cluster between-period ICC. This method has been evaluated by both Turner et al. and Forbes et al. and performs well[3, 4].

**3. Systematic review methods**

The protocol for the review has been published[12]. We provide a brief overview of the methods, along with deviations from the planned methods.

*3.1. Literature search*

The following sources were searched (to December 2014) for CRXO trials: MEDLINE, PubMed, EMBASE and CINAHL Plus. In addition, CRXO methodology articles were searched to identify further references to CRXO trials. We searched for methodology articles in PubMed using the following search strategy: ((cluster[tiab] AND cross\*over[tiab]) OR cluster-crossover[tiab]) AND (method\*[tiab] OR design [tiab] OR calcul\*[tiab] OR analy\*[tiab]). A citation search of all identified methodology articles was performed in Web of Science. Finally the references of all eligible articles were screened for CRXO trials. No restriction was applied to the publication date.

*3.2. Trial inclusion criteria*

Trials and protocols for trials that met the following inclusion criteria were included in the review: the trial was undertaken in humans; the trial was reported in English; the allocation of the intervention was to clusters of individuals rather than individuals themselves; each cluster received each intervention in a sequence over time (conventional crossover design), or at least some clusters crossed over from one intervention to another (such as two- treatment-four-sequence designs AA, AB, BA, and BB); at least some clusters crossed each way between at least two interventions (e.g. one cluster received AB and one cluster received BA); and the intervention given in the one period was not deliberately intended by design to affect individuals in subsequent periods (e.g. interventions intended to change the prescribing behaviour of health care provider, where patients form the cluster). Two criteria were added to the planned criteria; see Table S1 for further details on changes to the published protocol.

*3.3. Selection of trials for inclusion in the review*

One author (SA) screened all titles and abstracts using the predefined eligibility criteria and 50% of the titles and abstracts were screened independently by at least one co-author. Full articles were then screened by one author (SA) using the predefined eligibility criteria. All eligible articles were double screened along with 20% of articles that were initially determined to be ineligible by SA. Differences were resolved by discussion or by referral to a third author. No ineligible articles were subsequently found to be eligible.

*3.4. Data extraction*

The data extraction form incorporated items from the CONSORT extension to cluster randomised trials [1] and systematic reviews of the design, reporting and methodological aspects of stepped wedge [17], individual crossover [18], and parallel group cluster randomised trials [19, 20] The data extraction form was piloted on five trials by each author. This resulted in modifications and clarifications to the form (see Additional file 1). Data were entered into a database (Microsoft Access 2010, Redmond, Washington, USA). The extracted information included identification of the design in the title or abstract, justification for using the design, acknowledgement of the underlying assumptions of the design, demographic details (country, setting, unit of clustering, type of intervention, and control), design characteristics, methods used in the trial (recruitment, randomisation, allocation, and blinding), reporting of baseline characteristics of the trial design, and statistical analysis (methods to estimate intervention effects and adjustment for covariates). The extracted design characteristics included: number of clusters, number of periods, number of cluster-periods (clusters × periods), number of individuals in the trial, number of interventions and the allocation of interventions to cluster-periods, the variability of the number of individuals between cluster-periods, the reported measure of similarity between the outcomes of individuals within a cluster within a given period, and the reported measure of similarity between outcomes of individuals within a cluster between different periods.

We wished to collect information on the range of statistical methods used within the trials, so implemented the following process to select outcomes and their associated statistical methods. We collected information on the primary outcome, where we defined primary using the following hierarchy: the first eligible primary outcome in the protocol document or first published paper for the study if there is no protocol document; the outcome used for the sample size calculation; or the first outcome listed in the methods section of the abstract. We then collected information on multiple secondary outcomes, selected using the following process: the outcome was reported in the abstract and of a different data type to the primary outcome; the outcome was reported in the abstract and of the same data type as the previously included outcomes but analysed by a different method; the outcome was reported in the article and of different data type or analysed by a different method as the previously included outcomes.

One author (SA) extracted data from all trials, and data from 20% of the trials was independently double data extracted by the co-authors. Three of the five authors (SA, JM, AF) reviewed the discrepancies arising from the double data extraction, and discussed processes for further reviewing items where there was inconsistency. We re-reviewed the following items for all trials: contamination as a justification for using the design; use of a washout period; blinding of the deliverers of the intervention; mean cluster-period size. The sample size outcome scale was reviewed in all trials that used a count analysis. The scale of the outcome measure used in the sample size calculation was reviewed in all trials that were initially classified as count or binary. Any binary outcome that can occur multiple times per person in the time period of measurement was classified as a count outcome. We classified binary outcomes that are associated with a time period separately to binary outcomes measured over a fixed time period, since the statistical and sample size issues for these outcomes need to be considered separately. SA reviewed the following fields again in ten randomly selected trials: method of recruitment, allocation and blinding; reporting of baseline characteristics; sample size outcome scale; and, use of covariates in analysis. AF and JM reviewed the mean cluster-period size again in 20 randomly selected trials.

*3.5. Defining the appropriateness of sample size and analysis methods used in the CRXO trials*

We classified the sample size and analysis methods in each trial as either ‘potentially appropriate’ or ‘inappropriate’. Given there is limited methodological research investigating the performance of sample size and analysis methods (Section 2), it was not possible to classify the methods as definitely ‘appropriate’. We attempted to replicate the sample size calculation to assist in classifying the sample size methodology. Any trial that reported a method which attempted to adjust for both the cluster randomisation and multiple period design aspects, or equivalently the within-cluster within-period and within-cluster between-period ICCs, was considered to use ‘potentially appropriate’ methodology. Methods that aggregated the data in each cluster-period were judged to have accounted for the cluster randomisation design aspect. Trials using either GEE or GLM methodology that reported applying robust standard errors were considered ‘potentially appropriate’ if the number of clusters in a trial was at least 30, based on recommendations for parallel group cluster randomised trials (Hayes, 2009; pg 223) [14]. However, the performance of robust standard errors has not been extensively studied in CRXO trials, and in reality a higher number of clusters may be required (Morgan et al., unpublished result) [7].. Trials that reported using the statistical packages R (R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. ) or SAS (Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.) to fit GEEs were assumed to have applied robust standard errors, since robust standard errors are fitted by default in these packages. Methods that did not adjust for both the cluster randomisation and multiple period design aspects, but explicitly stated that the within-cluster within-period and within-cluster between-period ICCs were assumed to be equal were considered also to be ‘potentially appropriate’.

*3.6. Analysis*

We present descriptive summary statistics using frequencies and percentages of responses to categorical data. Free text was classified and frequencies and percentages of the categories are presented. For continuous data the range and mean/SD or median/IQR are presented as appropriate. The individual trial data can be made available on request to the corresponding author.

**4. Results**

*4.1.* *Results of the search*

Figure 1 shows the flow diagram of the CRXO trial selection process. Of the 3425 records identified through database searching, 170 were duplicates and 3046 were ineligible based on screening of abstracts, leaving 209 full-text articles to assess for eligibility. Of these 209 articles, 98 were assessed as eligible. A further four articles were identified through the methodology article reference and citation search, and four articles from the references of eligible articles. In total 106 articles from 91 unique trials were included in the review (see Additional file 2). Seventy-nine trials had only one associated article (eight of which were protocols), nine trials had two associated articles, and three trials had three associated articles.

*4.2. Characteristics of the trials*

The earliest identified trial was published in 1974. Half of the 91 trials were published after 2006, with nine trials published in 2014 (Figure 2). Most trials were conducted in a developed country (n=86, 95%), and were conducted within only one country (n=86, 95%) (Table S2).

The types of clusters varied. Almost half of the trials (n=45, 49%) randomised hospitals or wards within hospitals, and of these, 19 (21%) randomised intensive care units. Thirteen trials (14%) each randomised individual health care providers and schools or classes (Table S3).

The trials investigated a wide range of diseases and conditions, and health care delivery models. Twenty trials (22%) investigated infection control, 11 (12%) investigated infectious diseases, 11 (12%) investigated cardiovascular disease, and ten (11%) examined the delivery of health services (Table 1).

*4.3. Design characteristics of the trials*

The most common trial design included two interventions (n=81, 89%) (Table 2). In 49 (54%) trials, the interventions were delivered directly to the individuals within the clusters. In 22 (24%) trials, the intervention was targeted at the health care provider rather than the individuals under their care, and in 14 (15%) trials the intervention was targeted at the organisation of the health care provider or health service delivery (Table 1).

Trials had a median of nine clusters (IQR: 4 – 21, range: 2 - 268) and had a median cluster-period size of 30 (IQR: 14 – 77, range: 2 - 1319). The majority of trials (69%) used two periods. Trials randomising hospitals used fewer clusters (median: 6, IQR: 2 -10, range: 2 - 46), but were larger in size (median cluster-period size: 57, IQR: 21 – 197, range: 5 - 1319) than trials randomising schools and health care providers. The same participants were included in all periods in 27 (30%) trials, with this occurring more commonly in trials randomising schools (Table 2).

In 45 (49%) trials a washout period was either incorporated into the design or the reason for not using one was explained. Trials randomising health care providers used a washout less frequently (3/13, 23%) than trials randomising hospitals (27/45, 60%) or schools (7/13, 54%) (Table 2).

Of the 91 included trials, 76 (84%) trials stated that clusters were assigned to intervention sequences at random. Only 35 (38%) of these trials provided sufficient detail to replicate the randomisation. Thirty trials (33%) used restricted randomisation to balance covariates between the intervention sequences. Consent to participate was sought from the individuals or those acting on their behalf in 32 trials (35%). In 30 (33%) trials consent to participate was obtained at the cluster level, while in 24 (26%) trials it was unclear how consent was sought. (Table 1).

*4.4. Justification for design*

Only 25 (27%) trials provided justification for both the cluster randomisation and crossover aspects of the design, while 42 (46%) trials provided no justification for either design aspect (Table 1). A justification was provided for the cluster randomisation aspect in 36 trials (40%) and a justification for the crossover aspect was provided in 38 trials (42%). Justification for cluster randomisation was given in only three of 13 (23%) of the trials that randomised health care providers, where interventions were primarily targeted at the cluster-level (nine of 13 trials). In contrast justification for cluster randomisation was given in 22 (49%) of hospital trials, where the interventions were more frequently targeted at the individual level (62% of hospital trials) (Table S4).

The main reasons cited for cluster randomisation (n=36) were to avoid contamination between individuals within the cluster by either the individuals themselves (11 trials, 31%) or by those delivering the intervention (11 trials, 31%), and because it wasn't practical to individually randomise (11 trials, 31%). The main reason cited for crossing over the intervention within a cluster (n=38) was to attempt to eliminate differences in cluster-level characteristics between clusters (27 trials, 71% of reasons cited) (Table S4).

*4.5. Statistical methods for sample size estimation*

Of the 91 trials, 53 provided some detail of a sample size calculation, 35 (38%) did not report a sample size calculation. None of the trials reported using unequal cluster-period sizes (Table 3).

Only nine trials (10%) used methods that appropriately accounted for both the within-cluster within-period and within-cluster between-period ICCs. Eleven trials (12%) used a method appropriate for a parallel group cluster randomised trial, therefore ignoring the crossover aspect of the design. The remaining 33 trials either assumed the observations were independent (31 trials), or the reporting of sample size methodology was insufficient to make any assessment (two trials).

Trials that used a sample size calculation for a parallel group cluster randomised design almost always reported the value used to account for the non-independence of outcomes within each cluster (10 out of 11 trials). Seven of these values were based on a best guess, two were taken from published research, and one trial quoted the maximum correlation between clusters that would guarantee 80% power for their fixed sample size. Of the nine trials accounting for the CRXO design, five reported the values used to account for the non-independence of outcomes within and between periods within a cluster. In a further one trial: values were not reported; the within-cluster within-period ICC and within-cluster between-period ICC were assumed to cancel such that the sample size calculation assumed the outcomes were independent; the estimated reduction in power through simulation was reported, without reporting the values used within the simulation; and the sample size was inflated by a best guess value.

*4.6. Statistical methods for data analysis*

Across the 91 trials, 175 outcomes (median=2 outcomes per trial, IQR=1 to 2) met our inclusion criteria, from which we assessed the associated analytical methods. We excluded 36 of the analytical methods from further assessment because: the level of analysis was not clear (n=12); the intervention effect was estimated separately in each cluster or the intervention effect was estimated between clusters separately in each period (n=10); no comparison between intervention groups was made (n=9); no information on the method used for comparison was provided (n=3); or only descriptive statistics only were reported (n=2). Of the remaining 139 analyses, 127 (91%) were performed at the individual level and 12 (9%) were performed at the cluster level. Across the 139 analyses, we deemed 14 (10%) to be potentially appropriate. We now detail the methods used by the level at which the analysis was undertaken (Table 4).

Of the 12 cluster-level analyses, 10 (83%) used a method that accounted for the correlation between cluster-period summaries within each cluster and therefore appropriately accounted for both the cluster randomisation and crossover aspects of the design. In one analysis the methodology was judged to be inappropriate and the methodology in the remaining analysis was unclear.

In the 12 cluster-level analyses the observations were collapsed within cluster-periods to a cluster-period mean in five analyses, to a rate or count in another five analyses, and a log-incidence rate in one analysis. In the remaining analysis the expected rate in each cluster-period was obtained from a GLM fitted with a log-link function and individual-level covariates. The cluster-period summaries were compared using the following methods: paired t-test (seven analyses); permutation test (two analyses); and fitting a GLM with log-link function in two analyses, one applied robust standard errors to the estimate of the intervention effect and one included a random effect for cluster in the model. In the remaining analysis it was unclear whether a paired or unpaired t-test was used.

In contrast, 4 of the 127 individual-level analyses potentially accounted appropriately for both the cluster randomisation and crossover aspects of the design. Fifty-four analyses did not account for either the cluster randomisation or crossover aspects of the design. The cluster unit was accounted for in 52 analyses: 35 analyses fitted a GLM that included a term for the cluster (including eight fixed effects and 27 random effects, three also with robust standard errors); a GEE approach was used in 11 analyses (seven of which were judged to be using robust standard errors); four analyses used a Mantel Haenszel stratified chi-squared test; and two analyses fitted an ANOVA model with cluster as a fixed term. In the ten analyses that applied robust standard errors, four included at least 30 clusters and were therefore classified as potentially appropriate. In 21 analyses insufficient information was provided to determine if either the cluster randomisation or crossover aspects were accounted for. No trials accounted for the crossover as recommended by Turner et al. and Parienti et al. which involved including cluster-by-period random effect terms[4, 5].

*4.7. Reporting and handling of missing data*

Of 64 (70%) trials that reported missing data, 8 (13%) reported using a method to handle the missing data, including: use of random effects models, adjusted for covariates believed to be associated with missingness; multiple imputation to replace missing outcomes and covariates; and testing the sensitivity of the results to the missing data by substituting the missing values with extreme values.

**5. Discussion**

We undertook a systematic review to assess the design and statistical methods used in CRXO trials. CRXO trials have become more common over time, and have been conducted within a variety of settings to assess a range of interventions. The methods employed for the sample size calculation and analysis suggest that there was limited understanding of the effect of the cluster randomisation and multiple period aspects of the design. There is a need for improved reporting of CRXO trials: justifications for using the design were rarely reported.

A key requirement of the CRXO design is that the intervention effect does not carry over from one period to the next [5]. This can be achieved by many methods, for example using different participants in each period; providing an adequate washout time between periods; and blinding of the trialists involved in the delivery of the intervention and collection of outcome data. While it is possible to undertake a statistical test of the interaction between treatment and period, it is not possible from this test to distinguish carryover effects from treatment by period interactions, in two-period two-intervention CRXO trials[21], and altering the analysis based on the results of this test leads to a biased estimate of the intervention effect and inflated type I error[22]. Regardless of the method used to reduce the risk of carryover, clear reporting of the method used is required to allow readers to assess the risk of bias to the intervention effect arising from the potential carryover.

Although the CRXO design is usually more efficient than parallel group cluster randomisation, an individually randomised trial is usually more efficient than CRXO[3]. In over half of the CRXO trials included in our review, the intervention was delivered to the individual participants, and therefore individual randomisation may have been possible. Careful consideration should always be given to whether a more complex design is necessary, particularly when the design requires more participants[1].

*5.1 Implications of the sample size and analysis methodology*

*5.1.1. Implications of statistical methods for sample size estimation*

Over half (58%) of the trials that provided a sample size estimate reported using a method designed for an individually randomised trial using simple randomisation. This approach is only appropriate for a CRXO trial when both the within-cluster within-period ICC and within-cluster between-period ICC are zero, or the within-cluster between-period ICC is equal to within-cluster within-period ICC. These are strong and optimistic assumptions[3], and if violated, would lead to an underpowered study.

In trials where the sample size was estimated for a parallel group cluster randomised design, the within-cluster between-period ICC is effectively treated as zero. This approach leads to a conservative sample size and as a result, more participants will be included in the trial than are needed to obtain the desired power.

Following a suggestion by Donner et. al, one trial assumed the within-cluster between-period ICC was half the within-cluster within-period ICC, which leads to a sample size estimate that is *larger* than the requirement for an individually randomised design but *smaller* than the requirement for an parallel group cluster randomised trial[13]. In the absence of a priori knowledge of the within-cluster between-period ICC, this may be reasonable approach.

Other methods used to account for the CRXO design included inflating the sample size by an arbitrary amount; performing a power calculation using numerical simulation with a model that included both ICCs; basing the estimate on a paired t-test of the cluster-level means; and inflating the estimate from an individual level crossover design with the same participants in both periods by an inflation factor for a parallel group cluster randomised design. The first three methods may be appropriate if representative values of the sample size parameters were used. Further research is required to assess the appropriateness of the last method.

*5.1.2. Implications of choice of statistical methods for cluster-level analyses*

We found that when data was aggregated in each cluster-period, the analyses used were usually appropriate. In addition, as noted by Turner et al, such cluster-level analyses are intuitive and easily understandable to health researchers[4]. Therefore, use of such an analysis approach would often seem reasonable.

*5.1.3. Implications of choice of statistical methods for individual-level analyses*

Analogous to the issues with sample size, individual-level analyses that assume observations within clusters and within cluster-periods are independent, or account only for cluster-level variation, can estimate standard errors that are too small. These analyses can result in inflated type I error rates, and potentially lead to false positive claims regarding the effectiveness of interventions.

*5.1.4. Influence of statistical methodology articles on subsequent CRXO trials*

The influence of statistical methodology articles on the use of appropriate sample size and analysis methods in CRXO trials would seem limited. Methods to perform sample size calculations for CRXO trials have been available since 2004[10, 13]; yet, the number of papers including a sample size calculation has remained around 60% to 70% since 2000, and only two CRXO trials cited a methods paper for a CRXO sample size calculation. Methods for analysing binary data in dental split-mouth trials have been available since 2004[13], and in the context of CRXO trials, since 2007[5]. Only four CRXO trials cited a methods paper for analysing CRXO trials, all citing Turner et al[4]. It is unclear whether trialists recognise the need to use specialist methods when designing and analysing these trials. Regardless, however, there is a clear need for development of accessible guidance for health researchers for the design, conduct, and analysis of CRXO trials.

*5.2. Strengths and limitations*

A potential limitation of our review was our ability to locate all CRXO trials. Locating CRXO trials is difficult since there is no validated search strategy, and the language used to describe the design is inconsistent. Further, many trialists may be unaware that they have used a CRXO design, and so fail to use key words in the abstract that describe the clustering or crossover aspects. To optimise our yield of CRXO trials, we used a broad search strategy, searched references of all eligible CRXO trials, and undertook citation searches to CRXO trial methodology articles. Although our yield of CRXO trials may be incomplete, it represents the most comprehensive review of this trial design to date. Further, it may be argued that CRXO trials that are better reported, and thus easier to locate, are also more likely to use appropriate statistical methods. Therefore, results from our review may present an optimistic view of the design and statistical methods used in CRXO trials.

We maintained consistency in the review by having one author perform all screening and data extraction, and verified the results by having a subsample of all reviewed trials independently assessed by at least one other reviewer.

**6. Conclusions**

The CRXO design has been used in a wide range of settings for the past four decades. However, the statistical methods used in the sample size determination and analysis rarely account appropriately for the design aspects. The justifications for using the design are rarely reported. It is unclear whether trialists recognise the need for specialist methods in designing and analysing these trials. There is an urgent need for accessible guidance for health researchers on the design, conduct, analysis, and reporting of the CRXO design.

**Acknowledgments**

We acknowledge Ross Dunn for his assistance in the development of the Access database; and Steve McDonald for his assistance in the development of the search strategies. We thank Monica Taljaard for providing an unpublished data extraction form as a starting point for our data extraction form, particularly for the population details and intervention type sections.

S.J.A. was supported in part by a Monash University Graduate Scholarship and a National Health and Medical Research Council of Australia Centre of Research Excellence grant (1035261) to the Victorian Centre for Biostatistics (ViCBiostat).

K.E.M. was funded through a National Institute for Health Research (NIHR) research methods fellowship (MET-12-16).

J.E.M. was supported by a National Health and Medical Research Council (NHMRC) Australian Public Health Fellowship (1072366).

**References**

1. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C: **Consort 2010 statement: extension to cluster randomised trials**. *Bmj* 2012, **345**:e5661.

2. Bellomo R, Forbes A, Akram M, Bailey M, Pilcher DV, Cooper DJ: **Why we must cluster and cross over**. *Critical care and resuscitation : journal of the Australasian Academy of Critical Care Medicine* 2013, **15**(3):155-157.

3. Forbes AB, Akram M, Pilcher D, Cooper J, Bellomo R: **Cluster randomised crossover trials with binary data and unbalanced cluster sizes: application to studies of near-universal interventions in intensive care**. *Clin Trials* 2015, **12**(1):34-44.

4. Turner RM, White IR, Croudace T, Group PIPS: **Analysis of cluster randomized cross-over trial data: a comparison of methods**. *Statistics in medicine* 2007, **26**(2):274-289.

5. Parienti JJ, Kuss O: **Cluster-crossover design: a method for limiting clusters level effect in community-intervention studies**. *Contemp Clin Trials* 2007, **28**(3):316-323.

6. Rietbergen C, Moerbeek M: **The Design of Cluster Randomized Crossover Trials**. *Journal of Educational and Behavioral Statistics* 2011, **36**(4):472-490.

7. Eldridge S, Kerry S: **A Practical Guide to Cluster Randomised Trials in Health Services Research**: John Wiley & Sons, Ltd; 2012.

8. Louis TA, Lavori PW, Bailar JC, 3rd, Polansky M: **Crossover and self-controlled designs in clinical research**. *The New England journal of medicine* 1984, **310**(1):24-31.

9. Hooper R, Bourke L: **Cluster randomised trials with repeated cross sections: alternatives to parallel group designs**. *Bmj* 2015, **350:h2925**

10. Giraudeau B, Ravaud P, Donner A: **Sample size calculation for cluster randomized cross-over trials**. *Stat Med* 2008, **27**(27):5578-5585.

11. Reich NGM, Aaron M: **Improving efficiency in cluster-randomized study design and implementation: taking advantage of a crossover**. *Open Access Journal of Clinical Trials* 2014, **6**:11-15.

12. Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McDonald S, McKenzie JE: **The use of the cluster randomized crossover design in clinical trials: protocol for a systematic review**. *Systematic reviews* 2014, **3**:86.

13. Donner A, Klar N, Zou G: **Methods for the statistical analysis of binary data in split-cluster designs**. *Biometrics* 2004, **60**(4):919-925.

14. Hayes RJ, Moulton LH: **Cluster Randomised Trials**: Chapman and Hall/CRC; 2009.

15. Liang K-Y, Zeger SL: **Longitudinal data analysis using generalized linear models**. *Biometrika* 1986, **73**(1):13-22.

16. Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala SI, Wolfson M: **A comparison of two bias-corrected covariance estimators for generalized estimating equations**. *Biometrics* 2007, **63**(3):935-941.

17. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ: **Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation**. *J Clin Epidemiol* 2011, **64**(9):936-948.

18. Mills EJ, Chan AW, Wu P, Vail A, Guyatt GH, Altman DG: **Design, analysis, and presentation of crossover trials**. *Trials* 2009, **10**:27.

19. Rutterford C, Taljaard M, Dixon S, Copas A, Eldridge S: **Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: a review**. *J Clin Epidemiol* 2015, **68**(6):716-723.

20. Ivers NM, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, Skea Z, Brehaut JC, Boruch RF, Eccles MP *et al*: **Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8**. *Bmj* 2011, **343**:d5886.

21. Jones B, Kenwood MG: **Design and Analysis of Cross-Over Trials**, 2nd edn. London: Chapman & Hall/CRC; 2003.

22. Freeman PR: **The performance of the two-stage analysis of two-treatment, two-period crossover trials**. *Stat Med* 1989, **8**(12):1421-1432.

**Main Article Figures and Tables**

Figure 1: Flow of articles through the systematic review

Figure 2: Number of cluster randomised crossover publications per five year period.

Table 1: Characteristics of the cluster randomised crossover trials.

Table 2: Design characteristics of the cluster randomised crossover trials by type of cluster randomised

Table 3: Sample size methods used in the cluster randomised crossover trials

Table 4: Statistical analysis methods used in the cluster randomised crossover trials

**Supplementary Tables**

Table S1: Deviations from protocol methods

Table S2: Country where the trial was conducted

Table S3: Type of randomised cluster

Table S4: Justifications for cluster randomised crossover designs by type of cluster randomised

**Additional files**

Additional file 1: Data extraction form

Additional file 2: Reference list of trials included in systematic review

3425 records identified through database searching:

* Medline (n=2443)
* PubMed (n=33)
* EMBASE (n=782)
* CINAHL (n=167)

3046 records excluded after abstract screen due to:

* Article not in English (n=3)
* Outcomes not measured on humans or article is not a clinical trial (n=532)
* Trial design (n=2507)
  + Individually randomised (n=1627)
  + No crossover (n=102)
  + Individually randomised and no crossover (n=778)
* Each intervention given concurrently to different parts of the body (n=4)

3255 records after duplicates removed:

* Medline (n=2392)
* PubMed (n=27)
* EMBASE (n=706)
* CINAHL (n=130)

110 full text articles excluded due to:

* Outcome not measured on humans (n=1)
* Intervention applied to manikin or case study (n=3)
* Individually randomised (n=53)
* Clusters do not crossover (n=10)
* All clusters receive the same sequence of interventions (n=8)
* Not all interventions crossover (n=13)
* Outcomes measured at completion of all periods (n=3)
* Delayed intervention; carry over intended between periods (n=15)
* Carry over suspected; intention to assess effect of order (n=3)
* Each subject in cluster receives intervention for a different length of time (n=1)

209 full text articles assessed for eligibility

4 full text articles identified through methodology reference and citation search

3 full text articles identified from references of included articles

106 eligible articles

91 trials included in the review

Include additional 6 papers they need to be checked

Citation search of (n=4) methods papers and the references within returned an additional 5 papers

Figure 1: Flow of articles through the systematic review.



Figure 2: Number of cluster randomised crossover publications per five year period.

Table 1: Characteristics of the cluster randomised crossover trials

|  |  |
| --- | --- |
| **Disease or domain under study** | **N (%)** |
| Infection control | 20 (22%) |
| Infectious disease | 11 (12%) |
| Cardiovascular disease | 11 (12%) |
| Health services delivery | 10 (11%) |
| General and public health | 6 (7%) |
| Medical training | 5 (5%) |
| Communication of health information | 4 (4%) |
| Pregnancy, childbirth and early childhood | 3 (3%) |
| Mental health and behavioural conditions | 3 (3%) |
| Respiratory disease | 3 (3%) |
| Blood sample contamination | 3 (3%) |
| Cognition | 3 (3%) |
| Central nervous system and musculoskeletal disease | 2 (2%) |
| Urogenital disease | 2 (2%) |
| Oral health | 2 (2%) |
| Nutritional and metabolic disorders | 1 (1%) |
| Digestive disorders | 1 (1%) |
| Pain management | 1 (1%) |
| **Type of intervention** |  |
| Intervention targeting the individual | 49 (54%) |
| Intervention targeting health care provider | 22 (24%) |
| Quality improvement intervention | 14 (15%) |
| Intervention resulting in change to the participant environment | 6 (7%) |
| **Justification for design** |  |
| Justification for both cluster randomisation & crossover | 25 (27%) |
| Justification for neither cluster randomisation or crossover | 42 (46%) |
| Justification for cluster randomisation | 36 (40%) |
| Justification for crossover | 38 (42%) |
| **Consent** |  |
| Individual or those acting on their behalf | 32 (35%) |
| Cluster | 30 (33%) |
| Opt out | 4 (4%) |
| Varied by site1 | 1 1%) |
| Unclear | 24 (26%) |
| **Type of consent given by cluster-level decision maker2 (n=34)** |  |
| Participation and data collection | 6 (18%) |
| Participation, with individual consent for data collection | 23 (68%) |
| Not stated who gave consent for data collection | 5 (15%) |
| **Was the randomisation sequence randomly generated?3** |  |
| No | 3 (3%) |
| Yes - Sufficient information to replicate | 35 (38%) |
| Yes - Insufficient information to replicate | 41 (45%) |
| Unclear | 12 (13%) |
| **Covariates were used in the randomisation** | 30 (33%) |

1Consent was sought from the individuals or was obtained at the cluster level, varying by randomisation site

2Includes consent given by cluster-level decision maker and opt out consent

3A classification of no indicates that the treatment sequences were intentionally assigned to each cluster. The randomisation procedure was judged to be insufficient if reported that the allocation of treatment sequences was randomised, but no further detail on the randomisation was reported. If further details (e.g. toss of coin, computer randomisation program) were provided the procedure was judged to be sufficient.

Table 2: Design characteristics of the cluster randomised crossover trials by type of cluster randomised

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **All cluster types** | **Type of cluster randomised** | | | |
|  | **Total**  **N = 91** | **Hospital**  **N = 45** | **School**  **N = 13** | **Health care provider**  **N =13** | **Other cluster types1**  **N = 20** |
| **n (%)** | **n (%)** | **n (%)** | **n (%)** | **n (%)** |
| **Number of interventions** | | | | | |
| **2** | 81 (89%) | 38 (84%) | 12 (92%) | 12 (92%) | 19 (95%) |
| **3** | 9 (10%) | 6 (13%) | 1 (8%) | 1 (8%) | 1 (5%) |
| **4** | 1 (1%) | 1 (2%) | 0 | 0 | 0 |
| **Number of clusters - Median [IQR]; range** | 9 [4 - 21];  2- 268 | 6 [2 - 10];  2 – 46 | 17 [9 - 22];  4 – 46 | 23 [18 - 34];  3 – 64 | 14 [5 - 40];  2 – 268 |
| *Unclear* | *4 (4%)* | *2 (4%)* | *2 (15%)* | *0* | *0* |
| **Number of periods3** |  |  |  |  |  |
| 2 | 58 (69%) | 27 (60%) | 10 (83%) | 8 (73%) | 13 (81%) |
| 3 | 9 (11%) | 7 (16%) | 1 (8%) | 1 (9%) | 0 |
| 4+ | 17 (20%) | 11 (24%) | 1 (8%) | 2 (18%) | 3 (19%) |
| *Unclear* | *7 (8%)* | *0* | *1 (8%)* | *2 (15%)* | *4 (20%)* |
| **Cluster-period size - Median [IQR]; Range** | 30 [14-77];  2 - 1319 | 57 [21-194];  5 - 1319 | 23 [17-43];  10 - 152 | 20 [10-56]:  2 - 82 | 12 [6-27];  2 -77 |
| *Unclear* | 21 (23%) | 9 (20%) | 1 (8%) | 2 (15%) | 8 (40%) |
| **Same participants in all periods2** | 27 (30%) | 3 (7%) | 12 (93%) | 3 (23%) | 9 (45%) |
| **Washout period or reason for not including washout period explained** | 45 (49%) | 27 (60%) | 7 (54%) | 3 (23%) | 8 (40%) |

IQR: Interquartile Range

1 Other cluster types include: Aged care facilities, dementia facilities, primary care practices, and outpatient facilities; households and geographic regions; and worksite departments, emergency responder teams, and individual patients (units receiving treatment were individual teeth or muscles).

2 Data was dichotomised into same participants in all periods or no participants in multiple periods and some, but not all participants, in multiple periods .

3 Percentages of non-missing data presented

Table 3: Sample size methods used in the cluster randomised crossover trials

|  |  |
| --- | --- |
| **Trial reported a sample size calculation?** | **N=91**  **n (%)** |
| No | 35 (38%) |
| No - justification for not reporting calculation provided | 3 (3%) |
| Yes – sufficient information to replicate calculation | 39 (43%) |
| Yes – insufficient information to replicate calculation | 14 (15%) |
| **Sample size methods** | **N=53**  **n (%)** |
| Method included covariates in the sample size calculation?1 | 1 (2%) |
| Use of unequal cluster-period sizes?2 | 0 |
|  | **N=91**  **n (%)** |
| **Methods appropriate for individually randomised parallel group design (**Outcomes assumed to be independent) | 31 (34%) |
| **Methods appropriate for parallel group cluster randomised design** | 11 (12%) |
| Cluster accounted for, method unclear | 1 |
| Sample size inflated by design effect to account for within-cluster correlation | 10 |
| **Methods appropriate for CRXO design** | 9 (10%) |
| Paired cluster-level means | 1 |
| Sample size inflated by design effect that accounted for within-cluster within-period ICC and within-cluster between-period ICC | 4 |
| Sample size for individual crossover design inflated by design effect to account for within-cluster ICC | 1 |
| Sample size for stepped wedge design with verification using simulation | 1 |
| Sample size inflated by a best guess to account for CRXO design | 1 |
| Sample size estimated using simulation for CRXO design | 1 |
| **Methods unclear** | 2 (2%) |
| **Trial reported parameters used to account for correlation between outcomes in sample size calculation** |  |
| Methods appropriate for parallel group cluster randomised design (n=11) | 10 (91%) |
| Methods appropriate for CRXO design (n=9) | 5 (56%) |
|  |  |

CRXO: cluster randomised crossover

ICC: Intracluster correlation

1 Compared to no and unclear combined

2 Compared to equal cluster-period sizes. Cluster-period sizes were judged to be equal if the sample size could be reproduced by assuming equal sizes

Table 4: Statistical analysis methods used in the cluster randomised crossover trials

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Method** | **N** | **Potentially appropriate?** | **Cluster-level paired test** | **Permutation test** | **Fixed cluster effect6** | **Random cluster effect6** | **GEE** | **Robust standard errors7** | **Method did not account for cluster or cluster-period effect** | **Unclear if method accounted for cluster or cluster-period effect** |
| **Cluster level (N=12)** |  | 10 (83%) |  |  |  |  |  |  |  |  |
| GLM1 | 2 | 1 |  |  | 0 | 1 |  | 1 | 0 | 0 |
| ANOVA3 | 8 | 7 | 7 |  | 0 | 0 |  | 0 | 0 | 1 |
| Non-parametric methods5 | 2 | 2 | 0 | 2 | 0 | 0 |  | 0 | 0 | 0 |
| **Individual level (N=127)** |  | 4 (3%) |  |  |  |  |  |  |  |  |
| GLM1 | 64 | 1 |  |  | 8 | 24 |  | 3 | 15 | 14 |
| GEE2 | 11 | 3 |  |  |  |  | 4 | 7 |  |  |
| ANOVA3 | 13 |  |  |  | 2 |  |  |  | 7 | 4 |
| Other parametric models4 | 19 |  |  |  | 4 |  |  |  | 13 | 2 |
| Non-parametric methods5 | 20 |  |  |  |  |  |  |  | 19 | 1 |

Grey cells in the table indicate that the methodology in the column is not applicable to the analysis method in the row

GLM: Generalised Linear Model

ANOVA: Analysis of Variance

GEE: Generalised Estimating Equation

1 Models including linear, logistic, poisson, binomial-identity link, ordinal, proportional hazards, time series regression

2 Links including normal-identity, binomial-logit, binomial-identity, poisson-log

3 Including ANOVA, repeated measures ANOVA, t-test

4 Including Chi-squared, Mantel Haenszel chi-squared, McNemars test, Wallenstein method

5 Including Kruskal-Wallis, Fisher's exact, Kaplan-Meier curve with log-rank test, Wilcoxon rank sum, exact test for incidence rates, permutation tests

6 Including trials where the either the cluster-period effect was not accounted for in the method or it was unclear if the cluster-period effect was accounted for

7 For GLM a random effect for cluster was also used

**Supplementary Tables**

**Table S1: Summary of changes in methods from protocol methods**

|  |  |  |
| --- | --- | --- |
| **Protocol method** | **Deviation from protocol method** | **Justification** |
| ***Inclusion criteria*** |  |  |
| The trial was undertaken in humans; the allocation of the intervention was to clusters of individuals rather than individuals themselves; each cluster received each intervention in a sequence over time (conventional crossover design), or at least some clusters crossed over from one intervention to another (such as two- treatment-four-sequence designs AA, AB, BA, and BB) | At least some clusters crossed each way between at least two interventions.  *Type of deviation*: Addition | The intervention sequence in a CRXO design is randomly ordered to control for period effects. Several extracted trials applied a pre-post design to all clusters or applied a pre-post design to some clusters and a control intervention to the remaining clusters in all periods. |
| The intervention given in the one period was not deliberately intended by design to affect individuals in subsequent periods.  *Type of deviation*: Addition | Several extracted trials used designs where it was intended that the interventions in each intervention sequence would be compared separately, rather than pooled, to evaluate any ordering effects. |
| ***Full text review*** |  |  |
| Two reviewers will assess the full text articles. | All eligible articles were double screened along with 20% of articles that were initially determined to be ineligible by SA  *Type of deviation*: Amendment | The number of located CRXO trials was much greater than we had anticipated and we did not have the resources available for all full text articles to be double screened. We therefore amended our process to reduce the number full text articles that were double screened, but placed greater emphasis in this process on making the correct decision regarding non-inclusion of ineligible trials. |
| ***Data extraction form*** | Changes to data extraction were made after further piloting of the original data extraction form. See additional file 1.  *Type of deviation:* Amendment | After piloting the original data extraction form, changes were required to improve the clarity of the questions and to ensure that all data was extracted as intended in the original data extraction form. |
| ***Data extraction*** |  |  |
| Two reviewers will independently extract data using an electronic data extraction form developed for this review | One author (SA) extracted data from all trials, and data from 20% of the trials was independently double extracted by the co-authors. Three of the five authors (SA, JM, AF) reviewed the discrepancies arising from the double data extraction, and discussed processes for further reviewing items where there was inconsistency. | The number of located CRXO trials was much greater than we had anticipated and we did not have the resources available for all articles to have double data extraction. |

**Table S2: Country where the trial was conducted**

|  |  |
| --- | --- |
| **Country** | **N = 91**  **n (%)** |
| USA | 30 (33%) |
| UK | 10 (11%) |
| The Netherlands | 9 (10%) |
| More than 1 country | 5 (5%) |
| Canada | 5 (5%) |
| Australia | 4 (4%) |
| France | 4 (4%) |
| China | 2 (2%) |
| Denmark | 2 (2%) |
| Germany | 2 (2%) |
| Sweden | 2 (2%) |
| Thailand | 2 (2%) |
| Austria | 1 (1%) |
| Belgium | 1 (1%) |
| Estonia | 1 (1%) |
| Finland | 1 (1%) |
| Greece | 1 (1%) |
| Kenya | 1 (1%) |
| New Zealand | 1 (1%) |
| Pakistan | 1 (1%) |
| South Korea | 1 (1%) |
| South Africa | 1 (1%) |
| Switzerland | 1 (1%) |
| Taiwan | 1 (1%) |
| Tanzania | 1 (1%) |
| Zambia | 1 (1%) |

**Table S3: Type of randomised cluster**

|  |  |
| --- | --- |
| **Randomised cluster type** | **N = 91**  **n (%)** |
| Hospital or ward | 45 (49%) |
| ICU | 19 |
| Other wards | 26 |
| Individual health care provider | 13 (14%) |
| School or class | 13 (14%) |
| Class or classroom | 7 |
| School | 5 |
| Group of students | 1 |
| Emergency medical team | 6 (7%) |
| Primary care practice | 4 (4%) |
| Individual (mouth, muscles) | 2 (2%) |
| Dementia unit or facility | 2 (2%) |
| Aged care facility | 2 (2%) |
| Community or geographical area | 1 (1%) |
| Household or family group | 1 (1%) |
| Workplace | 1 (1%) |
| Outpatient clinic | 1 (1%) |

ICU: Intensive care unit

**Table S4: Justifications for cluster randomised crossover designs by type of cluster randomised**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **All cluster types** | **Type of cluster randomised** | | | |
|  | **N total**  **N = 91** | **Hospital**  **N = 45** | **School**  **N = 13** | **Health care provider**  **N=13** | **Other1**  **N=20** |
| **Justifications** |  |  |  |  |  |
| Justification for cluster randomisation | 36 (40%) | 22 (49%) | 4 (31%) | 3 (23%) | 7 (35%) |
| Justification for cross-over | 38 (42%) | 22 (49%) | 5 (39%) | 7 (54%) | 4 (20%) |
| **Cluster randomisation** | **N = 36** | **N = 22** | **N = 4** | **N = 3** | **N= 7** |
| Contamination likely between participants at the level of person/people delivering the intervention | 11 (31%) | 4 (18%) | 1 (25%) | 3 (100%) | 3 (43%) |
| Contamination likely between participants in a cluster | 11 (31%) | 8 (36%) | 2 (50%) | 0 | 1 (14%) |
| Practical/ethical/cost/administrative difficulties with randomising at an individual level | 11 (31%) | 6 (27%) | 2 (50%) | 1 (33%) | 2 (29%) |
| Ensure the intervention is fully delivered | 5 (14%) | 3 (14%) | 0 | 0 | 2 (29%) |
| Reflective of how the intervention will be applied in practice | 4 (11%) | 3 (14%) | 0 | 0 | 1 (14%) |
| Control for cluster level variation or achieve a balance of cluster covariates across interventions | 3 (8%) | 1 (5%) | 1 (25%) | 0 | 1 (14%) |
| Intervention only acts at the cluster level, impossible to randomise individually | 2 (6%) | 2 (9%) | 0 | 0 | 0 |
| Maximise the number of participants | 1 (3%) | 1 (5%) | 0 | 0 | 0 |
| Ensure blinding of participants is possible | 1 (3%) | 1 (5%) | 0 | 0 | 0 |
| Accepted trial design for school based interventions | 1 (3%) | 0 | 1 (25%) | 0 | 0 |
| **Cross over** | **N = 38** | **N = 22** | **N = 5** | **N = 7** | **N = 4** |
| Clusters expected to have different characteristics or to account for cluster level confounding | 27 (71%) | 17 (77%) | 4 (80%) | 4 (57%) | 2 (50%) |
| Reduced efficiency in estimating the intervention effect due to a limited number of clusters available for inclusion in the trial | 5 (13%) | 2 (9%) | 1 (20%) | 1 (14%) | 1 (25%) |
| Increased efficiency to overcome loss of power through randomising in clusters | 4 (11%) | 2 (9%) | 2 (40%) | 0 | 0 |
| Control period to "act as own control" (with no further justification) | 5 (13%) | 4 (18%) | 0 | 1 (14%) | 0 |
| Allow for within cluster comparison (with no further justification) | 1 (3%) | 1 (5%) | 0 | 0 | 0 |
| Increase power or precision (no further justification) | 4 (11%) | 1 (5%) | 0 | 2 (29%) | 1 (25%) |
| Increase participation | 3 (8%) | 2 (10%) | 1 (20%) | 0 | 0 |
| Reduce bias as each cluster contributes the same number of participants in each period | 2 (5%) | 1 (5%) | 0 | 0 | 1 (25%) |
| Allay ethical concerns by ensuring the intervention is received by all clusters | 1 (3%) | 1 (5%) | 0 | 0 | 0 |
| Allow for blinding of the allocation sequence. Participants would not know when they were in the control period | 1 (3%) | 0 | 0 | 1 (14%) | 0 |
| Permit historical and concurrent controls for each cluster | 1 (3%) | 1 (5%) | 0 | 0 | 0 |

1 Other cluster types include: Aged care facilities, dementia facilities, primary care practices, and outpatient facilities; households and geographic regions; and worksite departments, emergency responder teams, and individual patients (units receiving treatment were individual teeth or muscles).

**Additional files**

Additional file 1: Data extraction form

**CRXO systematic review data extraction form**

**Date: 16 December 2014**

**Section 1: Study Identifiers**

|  |  |
| --- | --- |
| Study ID | (Autocompleted) |
| 1.1 Is the article a protocol paper? | 0=No, 1=Yes |
| 1.1b Are there any other papers associated with this study? List references | String |
| 1.2 First Author Surname | String |
| 1.3 Publication Year | Integer, 1946 to 2014 |
| 1.5 Reviewer's initials | Categorical |
| 1.6 Date of review | Date |
| 1.7 Notes | Not to be analysed |

The unit of analysis for the review is ***study or trial***, not article. In many cases a study will be split into multiple articles, i.e. a protocol or design article, an article reporting the primary outcome(s), and many other articles reporting secondary outcomes.

The primary outcome for the study will be defined from the following hierarchy:

* The first primary outcome in the protocol document or first published paper for the study if there is no protocol document.
* The outcome used for the sample size calculation.
* The first outcome listed in the methods section of the abstract.

**Section 2: Full Text Screening**

Does the article report a research trial that used or planned to use a CRXO design which incorporated the following design elements (*The article will only be marked for inclusion if “yes” is answered to all):*

|  |  |
| --- | --- |
| 2.1 Outcomes were measured on humans in a study or trial at either cluster or individual level. | (0=No, 1=Yes, 2=Unclear) |
| 2.2 Allocation of the intervention was at cluster level (The allocation does not have to be at random). | (0=No, 1=Yes, 2=Unclear) |
| 2.3 Each cluster received each intervention, or at least some clusters crossed over from one intervention to another (e.g. two-intervention-four-sequence designs AA, AB, BA, BB). | (0=No, 1=Yes, 2=Unclear) |
| 2.4 Each cluster received each intervention in a sequence over time, rather than concurrently in time. | (0=No, 1=Yes, 2=Unclear) |
| 2.5 At least some clusters crossed each way between at least two interventions (e.g. one cluster received AB and one cluster received BA) | (0=No, 1=Yes, 2=Unclear) |
| 2.6 The intervention given in the one period was not deliberately intended by design to affect individuals in subsequent periods (e.g. interventions intended to change the prescribing behaviour of health care provider, where patients form the cluster) | (0=No, 1=Yes, 2=Unclear) |

**Section 3: Title and Abstract**

**Rationale:** To assess how CRXO trials are identified in the title and abstract.

|  |  |
| --- | --- |
| **Title** |  |
| Is the trial identified as a **cluster** randomised **crossover** trial in title? (*note words ‘cluster’ and ‘crossover’ must be used, placement of hyphens is unimportant*) | (0=No, 1=Yes) |
| **Abstract** |  |
| Is the trial identified as a **cluster** randomised trial in abstract? (*note words ‘cluster’ must be used, placement of hyphens is unimportant*) | (0=No, 1=Yes) |
| If no, copy verbatim from abstract how the unit of randomisation was described in the abstract | TEXT |
| Is the trial identified as a **crossover** trial in abstract? (*note words ‘crossover’ must be used, placement of hyphens is unimportant*) | (0=No, 1=Yes) |
| If no, copy verbatim from abstract how the cross over of interventions was described in the abstract | TEXT |

**Section 4: Justification for the CRXO design**

**Rationale:** To understand why researchers are using the CRXO design and how they justify that decision.

Why was the CRXO design chosen? *For* ***each*** *of the following points enter (0=Not Discussed, 1=Yes, 2=Unclear). Select as many points as apply.*

|  |  |
| --- | --- |
| *Justification given by authors for* ***cluster*** *randomisation* |  |
| Intervention can **only** act at the cluster level, and therefore impossible to randomise individually *(e.g. if the intervention is an educational program for health care practitioners, or a program implemented publicly via radio or newspaper, the intervention will reach a group of people)*. | (0=Not Discussed, 1=Yes, 2=Unclear) |
| Practical/ethical/cost/administrative difficulties with randomising at an individual level. | (0=Not Discussed, 1=Yes, 2=Unclear) |
| Contamination likely between participants at the level of person/people delivering the intervention *(e.g. an educational intervention may be delivered to health care practitioners, and it may impossible for them to only apply the intervention to some individuals in their care and not others. Therefore contamination would occur in an individually randomised trial).* | (0=Not Discussed, 1=Yes, 2=Unclear) |
| Contamination likely between participants in a cluster *(e.g. a behavioural intervention may be delivered to schools, and it may be impossible to prevent primary caregivers from exchanging experiences, thereby contaminating each arm of the trial in an individually randomised trial).* | (0=Not Discussed, 1=Yes, 2=Unclear) |
| To ensure intervention is fully delivered *(if it is expected that compliance with the trial protocol will be reduced if members of a cluster were individually randomised)* | (0=Not Discussed, 1=Yes, 2=Unclear) |
| Outcome data only available at cluster level | (0=Not Discussed, 1=Yes, 2=Unclear) |
| Other, specify | TEXT |
|  |  |
| *Justification given by authors for* ***crossover*** *design* |  |
| Increased efficiency to overcome loss of power through randomising in clusters (*i.e. the authors specifically cite a reduction in precision/power due to cluster randomisation or the ‘design effect’ as the reason for the crossover element).* | (0=Not Discussed, 1=Yes, 2=Unclear) |
| Reduced efficiency in estimating the intervention effect due to a limited number of clusters available for inclusion in the trial *(i.e. the authors cite the limited number of clusters as the reason for the crossover element).* | (0=Not Discussed, 1=Yes, 2=Unclear) |
| Clusters are expected to have different characteristics from each other or to account for cluster level confounding (*i.e. the authors cite that they expect or wish to allow for clusters being very different in characteristics which might affect the outcome, and wish to crossover so that each cluster ‘acts as own control’)* | (0=Not Discussed, 1=Yes, 2=Unclear) |
| Other, specify | TEXT |

**Section 5: Trial Objectives**

**Rationale:** What are the levels of the primary objective being addressed with the CRXO trial?

|  |  |
| --- | --- |
| Copy **verbatim** objective or hypothesis from Introduction | TEXT |

**Section 6: Population Details**

**Rationale:** What settings/conditions are CRXO trials being used in?

|  |  |
| --- | --- |
| Disease or domain under study  (can select multiple) | 1. Cancer  2. Cardiovascular  3. Central nervous system/musculoskeletal  4. Digestive/endocrine  5. Nutritional and metabolic  6. Gynaecology  7. Pregnancy and birth and paediatrics  8. Infectious diseases  9. Mental health and behavioural conditions  10. Pathological conditions  11. Symptoms and signs  12. Respiratory disease  13. Urogenital  14. Blood and immune system  15. Ear and nose  16. Eye  17. General health / public health  18. Genetic disorders  19. Injuries  20. Accidents and wounds  21. Mouth and dental  22. Skin  23. Other |
| Country of trial (List all if 5 or less, otherwise state multinational) | TEXT |
| Setting (select one) | 1. Primary care practices/health care clinics  2. Communities/geographical areas  3. Households/families  4. Aged care facility  5. Hospital  6. Schools  7. Workplaces  8. Other,specify |
| List if setting is other | TEXT |
| Do the methods define the cluster unit? | (0=No, 1=Yes, 2=Description unclear) |
| Clusters receiving intervention (select one) | 1. Primary care practices (practice includes multiple health care professionals) 2. Individual health professional 3. Communities/Residential areas 4. Households/families 5. Hospital, specify unit/ward type 6. Nursing home/aged care 7. Schools 8. Worksites 9. Other, specify |
| If hospital, specify ward or unit | TEXT |
| If other, specify | TEXT |
| Additional comments about clusters receiving intervention | TEXT |

**Section 7: Study Design**

**Rationale:** This section is intended to capture both the key design features of the published CRXO trial and how the design features are reported.

*An answer of “yes” means that the design aspect could be reconstructed from the information provided in the article.*

|  |  |
| --- | --- |
| The number of participating clusters | Integer, 99 = Not reported |
| The number of periods | Integer, 99 = Not reported |
| The number of interventions |  |
| Intervention treatments (*active interventions*) | Integer, 99 = Not reported |
| Control treatments (*e.g. no treatment, usual care. Enter 0 if all interventions are active*) | Integer, 99 = Not reported |
| List the different unique intervention sequences, (*i.e. AB, BA; or AA, BB, AB, BA*.) | TEXT |
| Is a diagram included to describe design? | (0=No, 1=Yes-complete, 2=Yes-incomplete/unclear) |
| Is each period ***designed*** to include the same or different participants in each period? (*i.e. are measurements repeated or not repeated on participants?)* | (0=Same participants, 1=Different participants, 2=Mix of same and different participants, 3=Unclear) |
| Do the authors ***discuss*** how many interventions each ***participant*** can receive, i.e. if the participant can remain in cluster for longer than one period or could be included again in a later period? | (0=Not Discussed, 1=Yes, 2=Unclear, 3=NA) |
| If the study is designed to include different participants in each period, does the study design make it possible for participants to be included in more than one period? | (0=No, 1=Yes, 2=Unclear) |
| Describe why you think it is possible that participants are in more than one period | TEXT |
| Are there any other relevant design features that may lead to additional correlation within the outcomes? E.g. hierarchical designs where there is clustering at different levels; wards within hospitals, GPs within general practices. (*copy* ***verbatim*** *from text*) | TEXT |

**Section 8: Carry over**

**Rationale:** To describe whether the risk of carry over being is acknowledged and managed.

|  |  |
| --- | --- |
| Do the authors ***discuss*** the possibility of carry over of intervention effects between periods? | (0=Not Discussed, 1=Yes, 2=Unclear*),*  Page and paragraph number |
| Is a washout period included? | (0=No and absence not explained, 1=Yes or absence explained, 2 = Not clear) |
| Was carry over managed in any other way? | (0=None listed, 1=Yes, text copied below) |
| Copy verbatim other ways in which carry over was managed | TEXT |
| If carry over possible, assess the risk of carry over effects (select one) | 1=Unlikely  2=Possible  3=Likely  4=Unclear |
| Describe your rationale for the assessment of the risk of carry over in above question | TEXT |

**Section 9: Blinding, selection bias and consent**

**Rationale:** To understand how randomisation or allocation of interventions was performed, the risk of bias in CRXO trials, and the adequacy of reporting.

*An answer of “yes” means that the design aspect could be reconstructed from the information provided in the article.*

|  |  |
| --- | --- |
| **Allocation sequence** |  |
| Was the allocation sequence randomly generated?  *(Where random is taken to mean: random number table, computer random number generator, coin tossing, shuffling cards or envelopes, throwing a dice, drawing of lots, minimisation)* | (0=No, 1=Yes-Sufficient to replicate, 2=Yes-Insufficient to replicate, 3=Unclear) |
| Were any covariates used in the randomisation scheme? *(ie stratification, minimisation, matching based on one or more covariates?)* | (0=None listed, 1=Yes, 2=Unclear) |
| ***If yes, complete for each covariate used in sample size calculation*** |  |
| *Covariate name:*  *(seccova\_covname)* | *Covariate level (1=individual, 2=cluster, 3=both, 4=unclear):*  *(seccova\_level)* |
|  |  |
| **Selection bias** |  |
| *Research team* |  |
| Is the allocation sequence known to the people allocating the intervention sequence to the clusters? I.e. can particular intervention sequences be deliberately matched to the clusters? *(Allocation concealment)* | (0=No, 1=Yes, 2=Unclear) |
| Do the people recruiting/identifying ***participants*** know which intervention sequence has been assigned to the cluster? | (0=No, 1=Yes, 2=Unclear, 3=NA-All participants recruited/identified before cluster randomisation, 4=NA-no recruitment/identification takes place) |
| Can the people recruiting/identifying ***participants*** influence which people are recruited/identified for inclusion in the study? | (0=No, 1=Yes, 2=Unclear)  Provide text to justify judgement – *e.g. participants are identified systematically from administrative data so identifier cannot influence inclusion.* |
| Provide text to justify judgement | TEXT |
| *Individual participants* |  |
| Who provides consent for the individual ***participant*** to ***receive intervention***? | 1. Individual. Consent is given by individual prior to intervention.  2. Cluster level. Individual participant does not give consent for intervention and cannot opt out of intervention. Consent is given by cluster spokesperson.  3. Opt out. Individual participant does not give consent for intervention. Intervention will be given unless participant opts out of intervention. Consent is given by cluster spokesperson.  4. Delayed consent. Consent is obtained from individual or their next of kin to continue intervention, but intervention is initiated without individual consent. Initial consent is given by cluster spokesperson.  5. Other  6. Unclear |
| Provide details if 'other' or ‘unclear’ is selected | TEXT |
| If the individual participant (or other person on their behalf) provides consent, does the ***participant*** have knowledge of the intervention to be receive first, ***prior*** to consenting? I.e. can the participant choose to take part because of the intervention they will receive first? | (0=No/unlikely, 1=Yes/possible, 2=Unclear, 3=NA (if option 2)) |
| Is the intervention concealed to ***participants*** ***during*** the study? (*I.e. is the intervention blinded?*) | (0=No, 1=Yes, 2=Unclear) |
|  |  |
| **Consent for data collection** |  |
| If the individual (or person on their behalf) does not provide consent, does consent for data collection occur at the cluster level, individual level, or is not required? | (0=Not reported, 1=cluster, 2=individual, 3=not required, 4=Reported but unclear) |
|  |  |
| **Performance bias** |  |
| Was the intervention concealed at cluster level (*i.e. were health care professionals delivering intervention blind to the intervention*)? | (0=No, 1=Yes, 2=Unclear) |
|  |  |
| **Detection bias** |  |
| Were any outcomes reported by the patient or individual collected (e.g. pain, depression)? (*e.g. pain, depression)?* | (0=No, 1=Yes, 2=Unclear) |
| If yes, list patient or individual level reported outcomes | TEXT |
| Were any subjective outcomes (e.g. clinician rated depression, condition specific mortality) collected by study personnel, clinicians, or outcome assessors (i.e. not reported by the individual level participant)? | (0=No, 1=Yes, 2=Unclear) |
| If yes, list which outcomes you considered subjective | TEXT |
| If yes, were the study personnel, clinicians, or outcome assessors who were assessing the subjective outcomes blind to the intervention assignment? | (0=No, 1=Yes, 2=Unclear) |

**Section 10: Intervention**

**Rationale:** To describe the type of interventions being used in CRXO

Type of experimental intervention (select all that apply. 0=No, 1=Yes, 2 =Unclear)

|  |  |
| --- | --- |
| Educational interventions that are **targeted at health care professionals** (e.g., distribution of educational materials, outreach visits, audit and feedback) | (0=No, 1=Yes, 2 =Unclear) |
| Quality improvement interventions **targeted at the organisation of health care or health delivery service.** The intervention is a new method for delivering or organising an existing health care service (e.g. 2 week vs 4 week attending physician rotation with trainees (change in delivery of medical training), daily vs on demand radiographs for mechanically ventilated patients (change in delivery of routine procedure), rapid detection test vs culture test of screening for MRSA carriage on admission (change in method of performing routine screening), financial, shifting of professional roles, multi-disciplinary teams, integration of services, changes in setting or equipment.) Distinguish from interventions to assess the effectiveness of a method or service that is performed or delivered at the level of the health care provider. | (0=No, 1=Yes, 2 =Unclear) |
| Intervention is **targeted at the health care professional** to indirectly alter patient outcomes. The intervention involves a change in the practise or behaviour of the health care professional (e.g. gloving procedure during venipuncture, hand and forearm cleaning for surgery.) Distinguish from quality improvement interventions where the intervention is intended to change the process of delivery of an existing health care service. |  |
| Intervention is **targeted at the cluster environment** rather than individuals within the cluster. The intervention indirectly affects the individuals within the cluster through changes in the environment. E.g. ward cleaning regime, air quality maintenance. |  |
| Participant health promotion or educational intervention. **Intervention is delivered directly to individual**. (e.g., promotion of breastfeeding, smoking cessation intervention, decision aid, disease screening promotion) | (0=No, 1=Yes, 2 =Unclear) |
| **Intervention is delivered directly to participants in the cluster** (e.g. change in drug or drug regime within ward for a given health condition, music therapy, exercise program, vitamin supplementation, insecticide spraying)  Direct participant therapeutic intervention) Distinguish from **indirect** changes to patient therapies as a result of guideline adherence or changes at level of those delivering the intervention. | (0=No, 1=Yes, 2 =Unclear) |
| Other, specify | TEXT |
| Details of experimental intervention | TEXT |
| Control intervention (select one) | 1. Not reported 2. No active intervention, i.e. usual care 3. Minimal application for experimental intervention 4. Placebo intervention 5. Other active intervention 6. Other, specify |
| Specify if other: | TEXT |
| Details of control intervention | TEXT |

**Section 11: Sample size**

**Rationale:** To assess how sample size calculations are being performed and justified

***Correlation terminology:***

Indiviudual i, Cluster j, Period k

Within-cluster within-period correlation: Corr(y\_ijk, y\_i'jk)

Within-cluster between-period correlation: Corr(y\_ijk, y\_i'jk')

*For the questions which ask for a justification, these are yes/no questions, either a justification was provided or it was not. However the “unclear” option remains because circumstances may arise where it isn’t clear if the question applies.*

|  |  |
| --- | --- |
| Was a sample size/power calculation presented? | (0=No, 1=Yes-Sufficient to replicate, 2=Yes-Insufficient to be reproduced, 3=Unclear 4=Reason given for no sample size calculation) |
| If a reference or method was provided for the sample size calculation, provide the reference or details of the method *(copy* ***verbatim*** *from article)* | TEXT |
| Was there a justification for number of periods? *Ie did the authors state why they used the number of periods.* | (0=No, 1=Yes, 2=Unclear, 3=Not justified but set by number available) |
| Was there a justification for number of clusters? *Ie did the authors state why they used the number of clusters, e.g. only 10 clusters available in region* | (0=No, 1=Yes, 2=Unclear, 3=Not justified but set by number available) |
| Was there a justification for number of participants per cluster? *Ie did the authors state why they used the number of participants, e.g. sample size calculation, all that were available in class.* | (0=No, 1=Yes, 2=Unclear, 3=Not justified but set by number available) |
| Which outcome was the sample size calculation based on? | TEXT |
| What was the scale of the outcome? | 1. Continuous  2. Binary  3. Categorical  4. Count  5. Time to event  6. Other, specify |
| If "other" provide details | TEXT |
| Were equal (as opposed to unequal) cluster sizes assumed in the calculation? | (0=Unequal, 1=Equal, 2=Unclear) |
| Was the ***within-cluster within-period*** clustering taken into consideration in the calculation? | (0=No, 1=Yes, 2=Unclear) |
| Please provide additional details about the cluster level clustering if you selected unclear or if you feel that clustering was adequately accounted for but "within-cluster within-period" does not capture the type of clustering | TEXT |
| If yes, what was the scale of the value? | 1=correlation  2=variance components  3=design effect  4=other, specify |
| If other, specify | TEXT |
| If yes, what was the value? | Float |
| If yes, what was the reference or source for the value for the ICC? (*e.g. pilot study, previous published research, best guess, unpublished research)* | TEXT |
| Was the within-cluster between-period clustering taken into consideration in the calculation? | (0=No, 1=Yes, 2=Unclear)  TEXT |
| Please provide additional details about the period level clustering if "within-cluster between-period" does not capture the type of clustering | TEXT |
| If yes, what was the scale of the value? | 1=correlation  2=variance  3=other, specify |
| If "other", please specify | TEXT |
| If yes, what was the value? | Float |
| If yes, what was the reference/source of the value for the ICC? (*e.g. pilot study, previous published research, best guess, unpublished research)* | TEXT |
| Were any covariates included in the sample size calculation? | (0=No, 1=Yes, 2=Unclear) |
| ***If yes, complete for each covariate used in sample size calculation*** |  |
| *Covariate name:*  *(seccovc\_covname)* | *Covariate level (1=individual, 2=cluster, 3=both, 4=unclear):*  *(seccovc\_level)* |
| Any additional comments? | TEXT |

**Section 12: Outcomes and Results**

**Rationale:** To describe the outcome measures being assessed with the CRXO design, and how they were being assessed.

Complete this table for each of the following:

* The primary outcome from the study (or outcome used in sample size or first outcome listed in article abstract methods or otherwise elsewhere in the abstract)
* The first secondary outcome that is reported in the abstract that is of a different data type to the primary outcome
* The first secondary outcome that is reported in the abstract that is of the same data type but analysed by a different method
* The first secondary outcome that is reported in the article that is of a different data type or analysed by a different method

|  |  |
| --- | --- |
| Specify outcome (*copy* ***verbatim*** *from text*) | TEXT |
| Classify how the outcome was identified from the study: | Primary outcome:  1. First primary outcome in the protocol document or published article  2. The outcome used for the sample size calculation  3. The first outcome listed in the article abstract  Secondary outcomes:  4. First outcome reported in abstract/protocol that is of a different data type to the primary outcome  5. First outcome reported in article that is of different data type or analysis |
| What type of data is the outcome? (select one) | 1=Continuous  2=Binary  3=Categorical  4=Count  5=Time to event  6=Other, specify |
| If “other” describe | TEXT |
| How was the statistical analysis concerning the intervention effect performed for the outcome? (*copy* ***verbatim*** *from text*) | TEXT |
| Was a justification given for the choice of analysis? *(E.g. Was a justification given for why they chose a multilevel individual level approach rather than a cluster level approach)* | (0=No, 1=Yes, 2=Unclear) |
| What justification was given for the choice of analysis? Copy verbatim: | TEXT |
| What references were provided for the statistical analysis and/or justification of the analysis? | TEXT |
|  |  |
| Was the within-cluster within-period clustering accounted for in the analysis of the outcome? | (0=No, 1=Yes, 2=Unclear) |
| Please provide additional details about the cluster level clustering if "within-cluster within-period" does not capture the type of clustering | TEXT |
| If reported, what was the scale of the clustering measure? | 1=ICC  2=variance components  3=coefficient of variation  4=not reported  5=other, specify |
| If "other" please specify | TEXT |
| If yes, what was the value of the clustering measure? | 99=Not reported |
| Was the within-cluster between-period clustering accounted for in the analysis of the outcome? | (0=No, 1=Yes, 2=Unclear) |
| Please provide additional details about the period level clustering if "within-cluster between-period" does not capture the type of clustering | TEXT |
| If reported, what was the scale of the clustering measure? | 1=ICC  2=variance components  3=coefficient of variation  4=not reported  5=other, specify |
| If "other" please specify |  |
| If yes, what was the value of the clustering measure? | 99=Not Reported |
| Were any other levels of clustering accounted for in the analysis? (*copy* ***verbatim*** *from text, e.g. repeated measurements on the same participant in a period*) | TEXT |
| Was the intervention effect adjusted for any covariates? | (0=No, 1=Yes, 2=Unclear) |
| ***Complete for each covariate used in analysis:*** |  |
| *Covariate name:*  *(seccovb\_covname)* | *Which level is the covariate measured at: (1=individual, 2=cluster, 3=both, 4=unclear):*  *(seccova\_measlevel)*  *Which level was the adjustment performed at:(1=individual, 2=cluster, 3=both, 4=unclear):*  *(seccova\_analysislevel)* |
| Provide details covariate adjustment *(copy* ***verbatim*** *from article)* | TEXT |
| Were the covariates used in randomisation included in the analysis? | (0=No, 1=Yes, 2=NA-none used in randomisation, 3=Unclear) |
| Any additional comments? | TEXT |

**Section 13: Baseline Characteristics (Table 1)**

**Rationale:** To describe how the baseline characteristics were summarised (Table 1).

In the following table, select all statements that apply.

|  |  |
| --- | --- |
| By intervention (i.e. separate summaries for the control and the intervention groups) *(NA if same participants are included in both interventions)* | (0=No, 1=Yes, 2=NA) |
| By period (i.e. separate summaries for each period for each intervention) *(NA if same participants are included in both periods)* | (0=No, 1=Yes, 2=NA) |
| By cluster (I.e. a separate summary for each cluster) | (0=No, 1=Yes) |
| By intervention sequence (I.e. a separate summary for each unique sequence of interventions) | (0=No, 1=Yes) |
| By total (I.e. a total summary for all participants) | (0=No, 1=Yes) |
| Other, specify | TEXT |

|  |  |
| --- | --- |
| Were the covariates used for randomisation reported in Table 1 (or a separate table)? | 0=No, 1=Yes-some, 2=Yes-all, 3=NA, 4=Unclear, 5=No table 1 |
| Were the covariates used for the sample size reported in Table 1 (or a separate table)? | 0=No, 1=Yes-some, 2=Yes-all, 3=NA, 4=Unclear, 5=No table 1 |
| Were the covariates used for analysis reported in Table 1 (or a separate table)? | 0=No, 1=Yes-some, 2=Yes-all, 3=NA, 4=Unclear, 5=No table 1 |

What were the sizes of the **analysed** clusters? If the values are not reported directly, but can be calculated from the supplied data, then perform the calculation and enter that value. I.e. 834 participants from 18 clusters gives a mean cluster size of 46.

|  |  |
| --- | --- |
| What was the mean number of participants in the cluster period? If multiple measurements are taken on each participant, report mean number of participants, not measurments. (99 = Not determinable) | Float |
| Was an indication provided for the variation in cluster size between clusters? | (0=No, 1=Yes, 2= No variation) |
| *If yes – copy text* ***verbatim*** *from article* | TEXT |
| Were any other summary statistics of the cluster sizes provided in the article? *(Copy text* ***verbatim*** *from article) (e.g. coefficient of variation, harmonic mean)* | TEXT |
| Was an indication provided for the variation in cluster size over time? *(e.g. between periods)* | (0=No, 1=Yes, 2 = No variation) |
| *If yes – copy text* ***verbatim*** *from article* | TEXT |

**Section: Missing data**

**Rationale:** To summarise whether missing data is being reported in the CRXO trials, and how the data is being accounted for in analyses

|  |  |
| --- | --- |
| Was missing data discussed in the article? | (0=No, 1=Yes, 2= Unlikely to be missing data) |
| How was missing data reported? (summarize, eg in text, diagram, poorly described, unclear): (*copy* ***verbatim*** *from text*) | TEXT |
| How did the authors account for missing data in the analysis? (*copy* ***verbatim*** *from text*) | TEXT |

Additional file 2: Reference list of trials included in systematic review

**Included Trials**

1. Lee N, Hui DSC, Zuo Z, Ngai KLK, Lui GCY, Wo SK, Tam WWS, Chan MCW, Wong BCK, Wong RYK, et al: **A prospective intervention study on higher-dose oseltamivir treatment in adults hospitalized with influenza a and B infections.** *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2013, **57:**1511-1519.

2. Adcock KG, Hogan SM, Elci OU, Mills KL: **Do Illustrations Improve Children's Comprehension of Assent Documents?** *The journal of pediatric pharmacology and therapeutics : JPPT : the official journal of PPAG* 2012, **17:**228-235.

3. Connolly SJ, Philippon F, Longtin Y, Casanova A, Birnie DH, Exner DV, Dorian P, Prakash R, Alings M, Krahn AD: **Randomized cluster crossover trials for reliable, efficient, comparative effectiveness testing: design of the Prevention of Arrhythmia Device Infection Trial (PADIT).** *The Canadian journal of cardiology* 2013, **29:**652-658.

4. Milstone AM, Elward A, Song X, Zerr DM, Orscheln R, Speck K, Obeng D, Reich NG, Coffin SE, Perl TM, Pediatric STSG: **Daily chlorhexidine bathing to reduce bacteraemia in critically ill children: a multicentre, cluster-randomised, crossover trial.** *Lancet* 2013, **381:**1099-1106.

5. Climo MW, Yokoe DS, Warren DK, Perl TM, Bolon M, Herwaldt LA, Weinstein RA, Sepkowitz KA, Jernigan JA, Sanogo K, Wong ES: **Effect of daily chlorhexidine bathing on hospital-acquired infection.** *The New England journal of medicine* 2013, **368:**533-542.

6. Washer LL, Chenoweth C, Kim H-W, Rogers MAM, Malani AN, Riddell Jt, Kuhn L, Noeyack B, Jr., Neusius H, Newton DW, et al: **Blood culture contamination: a randomized trial evaluating the comparative effectiveness of 3 skin antiseptic interventions.** *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 2013, **34:**15-21.

7. Wilder-Smith A, Byass P, Olanratmanee P, Maskhao P, Sringernyuang L, Logan JG, Lindsay SW, Banks S, Gubler D, Louis VR, et al: **The impact of insecticide-treated school uniforms on dengue infections in school-aged children: study protocol for a randomised controlled trial in Thailand.** *Trials* 2012, **13:**212.

8. Lucas BP, Trick WE, Evans AT, Mba B, Smith J, Das K, Clarke P, Varkey A, Mathew S, Weinstein RA: **Effects of 2- vs 4-week attending physician inpatient rotations on unplanned patient revisits, evaluations by trainees, and attending physician burnout: a randomized trial.** *JAMA : the journal of the American Medical Association* 2012, **308:**2199-2207.

9. Twardella D, Matzen W, Lahrz T, Burghardt R, Spegel H, Hendrowarsito L, Frenzel AC, Fromme H: **Effect of classroom air quality on students' concentration: results of a cluster-randomized cross-over experimental study.** *Indoor air* 2012, **22:**378-387.

10. Chant C, Mustard M, Thorpe KE, Friedrich JO: **Nurse- vs nomogram-directed glucose control in a cardiovascular intensive care unit.** *American journal of critical care : an official publication, American Association of Critical-Care Nurses* 2012, **21:**270-278.

11. Lin L-C, Huang Y-J, Watson R, Wu S-C, Lee Y-C: **Using a Montessori method to increase eating ability for institutionalised residents with dementia: a crossover design.** *Journal of clinical nursing* 2011, **20:**3092-3101.

12. Mubi M, Janson A, Warsame M, Martensson A, Kallander K, Petzold MG, Ngasala B, Maganga G, Gustafsson LL, Massele A, et al: **Malaria rapid testing by community health workers is effective and safe for targeting malaria treatment: randomised cross-over trial in Tanzania.** *PloS one* 2011, **6:**e19753.

13. Hill LJB, Williams JHG, Aucott L, Thomson J, Mon-Williams M: **How does exercise benefit performance on cognitive tests in primary-school pupils?** *Developmental medicine and child neurology* 2011, **53:**630-635.

14. Jongerden IP, Buiting AG, Leverstein-van Hall MA, Speelberg B, Zeidler S, Kesecioglu J, Bonten MJ: **Effect of open and closed endotracheal suctioning on cross-transmission with Gram-negative bacteria: a prospective crossover study.** *Critical care medicine* 2011, **39:**1313-1321.

15. Wilson APR, Smyth D, Moore G, Singleton J, Jackson R, Gant V, Jeanes A, Shaw S, James E, Cooper B, et al: **The impact of enhanced cleaning within the intensive care unit on contamination of the near-patient environment with hospital pathogens: a randomized crossover study in critical care units in two hospitals.** *Critical care medicine* 2011, **39:**651-658.

16. Kim N-H, Kim M, Lee S, Yun NR, Kim K-H, Park SW, Kim HB, Kim N-J, Kim E-C, Park WB, Oh M-D: **Effect of routine sterile gloving on contamination rates in blood culture: a cluster randomized trial.** *Annals of internal medicine* 2011, **154:**145-151.

17. Hill L, Williams JHG, Aucott L, Milne J, Thomson J, Greig J, Munro V, Mon-Williams M: **Exercising attention within the classroom.** *Developmental medicine and child neurology* 2010, **52:**929-934.

18. Nthumba PM, Stepita-Poenaru E, Poenaru D, Bird P, Allegranzi B, Pittet D, Harbarth S: **Cluster-randomized, crossover trial of the efficacy of plain soap and water versus alcohol-based rub for surgical hand preparation in a rural hospital in Kenya.** *The British journal of surgery* 2010, **97:**1621-1628.

19. Marks GB, Ezz W, Aust N, Toelle BG, Xuan W, Belousova E, Cosgrove C, Jalaludin B, Smith WT: **Respiratory health effects of exposure to low-NOx unflued gas heaters in the classroom: a double-blind, cluster-randomized, crossover study.** *Environmental health perspectives* 2010, **118:**1476-1482.

20. Parm U, Metsvaht T, Sepp E, Ilmoja ML, Pisarev H, Pauskar M, Lutsar I: **Impact of empiric antibiotic regimen on bowel colonization in neonates with suspected early onset sepsis.** *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 2010, **29:**807-816.

21. Nijssen S, Fluit A, van de Vijver D, Top J, Willems R, Bonten MJM: **Effects of reducing beta-lactam antibiotic pressure on intestinal colonization of antibiotic-resistant gram-negative bacteria.** *Intensive care medicine* 2010, **36:**512-519.

22. Gerald LB, Gerald JK, Zhang B, McClure LA, Bailey WC, Harrington KF: **Can a school-based hand hygiene program reduce asthma exacerbations among elementary school children?** *The Journal of allergy and clinical immunology* 2012, **130:**1317-1324.

23. Hejblum G, Chalumeau-Lemoine L, Ioos V, Boelle P-Y, Salomon L, Simon T, Vibert J-F, Guidet B: **Comparison of routine and on-demand prescription of chest radiographs in mechanically ventilated adults: a multicentre, cluster-randomised, two-period crossover study.** *Lancet* 2009, **374:**1687-1693.

24. Rozendaal FW, Spronk PE, Snellen FF, Schoen A, van Zanten ARH, Foudraine NA, Mulder PGH, Bakker J, Ulti Si: **Remifentanil-propofol analgo-sedation shortens duration of ventilation and length of ICU stay compared to a conventional regimen: a centre randomised, cross-over, open-label study in the Netherlands.** *Intensive care medicine* 2009, **35:**291-298.

25. de Smet AMGA, Kluytmans JAJW, Cooper BS, Mascini EM, Benus RFJ, van der Werf TS, van der Hoeven JG, Pickkers P, Bogaers-Hofman D, van der Meer NJM, et al: **Decontamination of the digestive tract and oropharynx in ICU patients.** *The New England journal of medicine* 2009, **360:**20-31.

26. Kelly P, Katubulushi M, Todd J, Banda R, Yambayamba V, Fwoloshi M, Zulu I, Kafwembe E, Yavwa F, Sanderson IR, Tomkins A: **Micronutrient supplementation has limited effects on intestinal infectious disease and mortality in a Zambian population of mixed HIV status: a cluster randomized trial.** *The American journal of clinical nutrition* 2008, **88:**1010-1017.

27. Jeyaratnam D, Whitty CJM, Phillips K, Liu D, Orezzi C, Ajoku U, French GL: **Impact of rapid screening tests on acquisition of meticillin resistant Staphylococcus aureus: cluster randomised crossover trial.** *BMJ (Clinical research ed)* 2008, **336:**927-930.

28. Harbarth S, Fankhauser C, Schrenzel J, Christenson J, Gervaz P, Bandiera-Clerc C, Renzi G, Vernaz N, Sax H, Pittet D: **Universal screening for methicillin-resistant Staphylococcus aureus at hospital admission and nosocomial infection in surgical patients.** *JAMA : the journal of the American Medical Association* 2008, **299:**1149-1157.

29. Tortolero SR, Markham CM, Addy RC, Baumler ER, Escobar-Chaves SL, Basen-Engquist KM, McKirahan NK, Parcel GS: **Safer choices 2: rationale, design issues, and baseline results in evaluating school-based health promotion for alternative school students.** *Contemporary clinical trials* 2008, **29:**70-82.

30. Bleasdale SC, Trick WE, Gonzalez IM, Lyles RD, Hayden MK, Weinstein RA: **Effectiveness of chlorhexidine bathing to reduce catheter-associated bloodstream infections in medical intensive care unit patients.** *Archives of internal medicine* 2007, **167:**2073-2079.

31. Glazebrook C, Marlow N, Israel C, Croudace T, Johnson S, White IR, Whitelaw A: **Randomised trial of a parenting intervention during neonatal intensive care.** *Archives of disease in childhood Fetal and neonatal edition* 2007, **92:**F438-443.

32. Sloane PD, Williams CS, Mitchell CM, Preisser JS, Wood W, Barrick AL, Hickman SE, Gill KS, Connell BR, Edinger J, Zimmerman S: **High-intensity environmental light in dementia: effect on sleep and activity.** *Journal of the American Geriatrics Society* 2007, **55:**1524-1533.

33. Vessey JA, Sherwood JJ, Warner D, Clark D: **Comparing hand washing to hand sanitizers in reducing elementary school students' absenteeism.** *Pediatric nursing* 2007, **33:**368-372.

34. Lucas BP, Asbury JK, Wang Y, Lee K, Kumapley R, Mba B, Borkowsky S, Asmar A: **Impact of a bedside procedure service on general medicine inpatients: A firm-based trial.** *Journal of hospital medicine : an official publication of the Society of Hospital Medicine* 2007, **2:**143-149.

35. Fowell A, Johnstone R, Finlay IG, Russell D, Russell IT: **Design of trials with dying patients: a feasibility study of cluster randomisation versus randomised consent.** *Palliative medicine* 2006, **20:**799-804.

36. Sjogren T, Nissinen KJ, Jarvenpaa SK, Ojanen MT, Vanharanta H, Malkia EA: **Effects of a physical exercise intervention on subjective physical well-being, psychosocial functioning and general well-being among office workers: a cluster randomized-controlled cross-over design.** *Scandinavian journal of medicine & science in sports* 2006, **16:**381-390.

37. Thomas B, Jones M, Johns P, Trauer T: **P.r.n. medication use in a psychiatric high-dependency unit following the introduction of a nurse-led activity programme.** *International journal of mental health nursing* 2006, **15:**266-271.

38. Hoeffer B, Talerico KA, Rasin J, Mitchell CM, Stewart BJ, McKenzie D, Barrick AL, Rader J, Sloane PD: **Assisting cognitively impaired nursing home residents with bathing: effects of two bathing interventions on caregiving.** *The Gerontologist* 2006, **46:**524-532.

39. Lang E, Afilalo M, Vandal AC, Boivin J-F, Xue X, Colacone A, Leger R, Shrier I, Rosenthal S: **Impact of an electronic link between the emergency department and family physicians: a randomized controlled trial.** *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 2006, **174:**313-318.

40. Larson EL, Cimiotti J, Haas J, Parides M, Nesin M, Della-Latta P, Saiman L: **Effect of antiseptic handwashing vs alcohol sanitizer on health care-associated infections in neonatal intensive care units.** *Archives of pediatrics & adolescent medicine* 2005, **159:**377-383.

41. Cook DA, Dupras DM, Thompson WG, Pankratz VS: **Web-based learning in residents' continuity clinics: a randomized, controlled trial.** *Academic medicine : journal of the Association of American Medical Colleges* 2005, **80:**90-97.

42. Vinson DC, Galliher JM, Reidinger C, Kappus JA: **Comfortably engaging: which approach to alcohol screening should we use?** *Annals of family medicine* 2004, **2:**398-404.

43. Lam WWT, Fielding R, Johnston JM, Tin KYK, Leung GM: **Identifying barriers to the adoption of evidence-based medicine practice in clinical clerks: a longitudinal focus group study.** *Medical education* 2004, **38:**987-997.

44. Morton JL, Schultz AA: **Healthy Hands: Use of alcohol gel as an adjunct to handwashing in elementary school children.** *The Journal of school nursing : the official publication of the National Association of School Nurses* 2004, **20:**161-167.

45. Parienti J-J, du Cheyron D, Ramakers M, Malbruny B, Leclercq R, Le Coutour X, Charbonneau P, Members of the NSG: **Alcoholic povidone-iodine to prevent central venous catheter colonization: A randomized unit-crossover study.** *Critical care medicine* 2004, **32:**708-713.

46. Vrentzos GE, Papadakis JA, Vardakis KE, Maliaraki N, Stilianou K, Arvanitis A, Sratigis S, Alivanis P, Margioris AN, Ganotakis ES: **Intravenous administration of vitamin B12 in the treatment of hyperhomocysteinemia associated with end-stage renal disease.** *Journal of nephrology* 2003, **16:**535-539.

47. van Alem AP, Vrenken RH, de Vos R, Tijssen JGP, Koster RW: **Use of automated external defibrillator by first responders in out of hospital cardiac arrest: prospective controlled trial.** *BMJ (Clinical research ed)* 2003, **327:**1312.

48. Detmar SB, Muller MJ, Schornagel JH, Wever LDV, Aaronson NK: **Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial.** *JAMA : the journal of the American Medical Association* 2002, **288:**3027-3034.

49. Parienti JJ, Thibon P, Heller R, Le Roux Y, von Theobald P, Bensadoun H, Bouvet A, Lemarchand F, Le Coutour X, Antisepsie Chirurgicale des mains Study G: **Hand-rubbing with an aqueous alcoholic solution vs traditional surgical hand-scrubbing and 30-day surgical site infection rates: a randomized equivalence study.** *JAMA : the journal of the American Medical Association* 2002, **288:**722-727.

50. Calfee DP, Farr BM: **Comparison of four antiseptic preparations for skin in the prevention of contamination of percutaneously drawn blood cultures: a randomized trial.** *Journal of clinical microbiology* 2002, **40:**1660-1665.

51. Herkner H, Eisenburger P, Havel C, Laggner AN: **Dashing with scooters to in-hospital emergencies: a randomised cross-over experiment.** *Resuscitation* 2002, **52:**293-296.

52. Chan L, MacLehose RF, Houck PM: **Impact of physician reminders on the use of influenza vaccinations: a randomized trial.** *Archives of physical medicine and rehabilitation* 2002, **83:**371-375.

53. Cibere J, Sibley JT, Haga M: **Rheumatologists' adherence to guidelines for misoprostol use in patients at high risk for nonsteroidal antiinflammatory drug gastropathy.** *The Journal of rheumatology* 2002, **29:**339-346.

54. Karchmer TB, Giannetta ET, Muto CA, Strain BA, Farr BM: **A randomized crossover study of silver-coated urinary catheters in hospitalized patients.** *Archives of internal medicine* 2000, **160:**3294-3298.

55. Steele DJ, Medder JD, Turner P: **A comparison of learning outcomes and attitudes in student- versus faculty-led problem-based learning: an experimental study.** *Medical education* 2000, **34:**23-29.

56. Dahler-Eriksen BS, Lauritzen T, Lassen JF, Lund ED, Brandslund I: **Near-patient test for C-reactive protein in general practice: assessment of clinical, organizational, and economic outcomes.** *Clinical chemistry* 1999, **45:**478-485.

57. Glazener CM, Ramsay CR, Campbell MK, Booth P, Duffty P, Lloyd DJ, McDonald A, Reid JA: **Neonatal examination and screening trial (NEST): a randomised, controlled, switchback trial of alternative policies for low risk infants.** *BMJ (Clinical research ed)* 1999, **318:**627-631.

58. Chavasse DC, Shier RP, Murphy OA, Huttly SR, Cousens SN, Akhtar T: **Impact of fly control on childhood diarrhoea in Pakistan: community-randomised trial.** *Lancet* 1999, **353:**22-25.

59. Bader HI, Boyd RL: **Comparative efficacy of a rotary and a sonic powered toothbrush on improving gingival health in treated adult periodontitis patients.** *American journal of dentistry* 1999, **12:**143-147.

60. Flanagan JR, Doebbeling BN, Dawson J, Beekmann S: **Randomized study of online vaccine reminders in adult primary care.** *Proceedings / AMIA Annual Symposium AMIA Symposium* 1999**:**755-759.

61. Desai J, Sharief M, Swash M: **Riluzole has no acute effect on motor unit parameters in ALS.** *Journal of the neurological sciences* 1998, **160 Suppl 1:**S69-72.

62. Jernigan JA, Siegman-Igra Y, Guerrant RC, Farr BM: **A randomized crossover study of disposable thermometers for prevention of Clostridium difficile and other nosocomial infections.** *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 1998, **19:**494-499.

63. Mendelson MH, Short LJ, Schechter CB, Meyers BR, Rodriguez M, Cohen S, Lozada J, DeCambre M, Hirschman SZ: **Study of a needleless intermittent intravenous-access system for peripheral infusions: analysis of staff, patient, and institutional outcomes.** *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 1998, **19:**401-406.

64. Borutta A: **Plaque removal efficacy of a newly developed powered toothbrush in the primary dentition of pre-school children.** *The Journal of clinical dentistry* 1997, **8:**151-155.

65. Barrett BJ, Parfrey PS, McDonald J, Haire R, Peachey G: **Concurrent utilization review and inappropriate hospital stay: evaluation of a program.** *Clinical and investigative medicine Medecine clinique et experimentale* 1996, **19:**28-35.

66. Skargren E, Oberg B: **Effects of an exercise program on musculoskeletal symptoms and physical capacity among nursing staff.** *Scandinavian journal of medicine & science in sports* 1996, **6:**122-130.

67. Schwab TM, Callaham ML, Madsen CD, Utecht TA: **A randomized clinical trial of active compression-decompression CPR vs standard CPR in out-of-hospital cardiac arrest in two cities.** *JAMA : the journal of the American Medical Association* 1995, **273:**1261-1268.

68. Nydahl M, Gustafsson IB, Ohrvall M, Vessby B: **Similar serum lipoprotein cholesterol concentrations in healthy subjects on diets enriched with rapeseed and with sunflower oil.** *European journal of clinical nutrition* 1994, **48:**128-137.

69. Evans RS, Classen DC, Pestotnik SL, Lundsgaarde HP, Burke JP: **Improving empiric antibiotic selection using computer decision support.** *Archives of internal medicine* 1994, **154:**878-884.

70. Kellermann AL, Hackman BB, Somes G, Kreth TK, Nail L, Dobyns P: **Impact of first-responder defibrillation in an urban emergency medical services system.** *JAMA : the journal of the American Medical Association* 1993, **270:**1708-1713.

71. Thamlikitkul V, Chokloikaew S, Tangtrakul T, Siripoonkiat P, Wongpreedee N, Danchaivijitr S: **Blood culture: comparison of outcomes between switch-needle and no-switch techniques.** *American journal of infection control* 1992, **20:**122-125.

72. Larson E, McGeer A, Quraishi ZA, Krenzischek D, Parsons BJ, Holdford J, Hierholzer WJ: **Effect of an automated sink on handwashing practices and attitudes in high-risk units.** *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 1991, **12:**422-428.

73. Anonymous: **Dry chemistry pathology in general practice. An assessment of utilisation, perceptions and impact on patient care. Non-Laboratory Pathology Working Party of the National Health Technology Advisory Panel.** *The Medical journal of Australia* 1991, **154:**511-518.

74. Epperson EL: **Efficacy of 0.9% sodium chloride injection with and without heparin for maintaining indwelling intermittent injection sites.** *Clinical pharmacy* 1984, **3:**626-629.

75. Barker KN, Pearson RE, Hepler CD, Smith WE, Pappas CA: **Effect of an automated bedside dispensing machine on medication errors.** *American journal of hospital pharmacy* 1984, **41:**1352-1358.

76. Hayes MJ, Morris GK, Hampton JR: **Comparison of mobilization after two and nine days in uncomplicated myocardial infarction.** *British medical journal* 1974, **3:**10-13.

77. Szyld E, Aguilar A, Musante GA, Vain N, Prudent L, Fabres J, Carlo WA: **Comparison of Devices for Newborn Ventilation in the Delivery Room.** *The Journal of pediatrics* 2014.

78. Guy RJ, Natoli L, Ward J, Causer L, Hengel B, Whiley D, Tabrizi SN, Donovan B, Fairley CK, Badman SB, et al: **A randomised trial of point-of-care tests for chlamydia and gonorrhoea infections in remote Aboriginal communities: Test, Treat ANd GO- the "TTANGO" trial protocol.** *BMC infectious diseases* 2013, **13:**485.

79. van Werkhoven CH, Postma DF, Oosterheert JJ, Bonten MJ: **Antibiotic treatment of moderate-severe community-acquired pneumonia: design and rationale of a multicentre cluster-randomised cross-over trial.** *The Netherlands journal of medicine* 2014, **72:**170-178.

80. Hardy K, Price C, Szczepura A, Gossain S, Davies R, Stallard N, Shabir S, McMurray C, Bradbury A, Hawkey PM: **Reduction in the rate of methicillin-resistant Staphylococcus aureus acquisition in surgical wards by rapid screening for colonization: a prospective, cross-over study.** *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2010, **16:**333-339.

81. Hostler D, Everson-Stewart S, Rea TD, Stiell IG, Callaway CW, Kudenchuk PJ, Sears GK, Emerson SS, Nichol G, Resuscitation Outcomes Consortium I: **Effect of real-time feedback during cardiopulmonary resuscitation outside hospital: prospective, cluster-randomised trial.** *Bmj* 2011, **342:**d512.

82. Scales DC, Dainty K, Hales B, Pinto R, Fowler RA, Adhikari NK, Zwarenstein M: **A multifaceted intervention for quality improvement in a network of intensive care units: a cluster randomized trial.** *Jama* 2011, **305:**363-372.

83. Stiell IG, Nichol G, Leroux BG, Rea TD, Ornato JP, Powell J, Christenson J, Callaway CW, Kudenchuk PJ, Aufderheide TP, et al: **Early versus later rhythm analysis in patients with out-of-hospital cardiac arrest.** *The New England journal of medicine* 2011, **365:**787-797.

84. Damsgaard CT, Dalskov SM, Laursen RP, Ritz C, Hjorth MF, Lauritzen L, Sorensen LB, Petersen RA, Andersen MR, Stender S, et al: **Provision of healthy school meals does not affect the metabolic syndrome score in 8-11-year-old children, but reduces cardiometabolic risk markers despite increasing waist circumference.** *The British journal of nutrition* 2014, **112:**1826-1836.

85. Berndt N, Bolman C, Froelicher ES, Mudde A, Candel M, de Vries H, Lechner L: **Effectiveness of a telephone delivered and a face-to-face delivered counseling intervention for smoking cessation in patients with coronary heart disease: a 6-month follow-up.** *Journal of behavioral medicine* 2014, **37:**709-724.

86. Oostdijk EAN, Kesecioglu J, Schultz MJ, Visser CE, de Jonge E, van Essen EHR, Bernards AT, Purmer I, Brimicombe R, Bergmans D, et al: **Effects of decontamination of the oropharynx and intestinal tract on antibiotic resistance in ICUs: a randomized clinical trial.** *Jama* 2014, **312:**1429-1437.

87. Reddy SK, Bailey MJ, Beasley RW, Bellomo R, Henderson SJ, Mackle DM, McArthur CJ, Mehrtens JE, Myburgh JA, McGuinness SP, et al: **A protocol for the 0.9% saline versus Plasma-Lyte 148 for intensive care fluid therapy (SPLIT) study.** *Critical care and resuscitation : journal of the Australasian Academy of Critical Care Medicine* 2014, **16:**274-279.

88. van Duijn PJ, Bonten MJ: **Antibiotic rotation strategies to reduce antimicrobial resistance in Gram-negative bacteria in European intensive care units: study protocol for a cluster-randomized crossover controlled trial.** *Trials* 2014, **15:**277.

89. Pourrat X, Roux C, Bouzige B, Garnier V, Develay A, Allenet B, Fraysse M, Halimi JM, Grassin J, Giraudeau B: **Impact of drug reconciliation at discharge and communication between hospital and community pharmacists on drug-related problems: study protocol for a randomized controlled trial.** *Trials* 2014, **15:**260.

90. Roisin S, Laurent C, Denis O, Dramaix M, Nonhoff C, Hallin M, Byl B, Struelens MJ: **Impact of rapid molecular screening at hospital admission on nosocomial transmission of methicillin-resistant staphylococcus aureus: cluster randomised trial.** *PloS one* 2014, **9:**e96310.

91. Surka S, Edirippulige S, Steyn K, Gaziano T, Puoane T, Levitt N: **Evaluating the use of mobile phone technology to enhance cardiovascular disease screening by community health workers.** *International Journal of Medical Informatics* 2014, **83:**648-654.