

Review Article

Cluster-Randomized Studies

Part 25 of a Series on Evaluating Scientific Publications

Eva Lorenz, Sascha Köpke, Holger Pfaff, Maria Blettner

Summary

Background: Cluster-randomized trials (CRT) are needed to compare interventions that are allocated to entire groups of subjects, rather than to individuals. Publications about CRT have become steadily more common over the past decade. Readers of such publications should be able to categorize and interpret the findings of CRT correctly while considering the methodological requirements applicable to this type of study.

Methods: This review is based on a selection of pertinent literature and on the authors' expertise. CRT-specific methodological aspects of the planning, performance, and interpretation of studies are discussed.

Results: Readers of publications on CRT should check whether due consideration has been given to correlations within and between the clusters during the planning of the study. These correlations enable the determination whether persons within a cluster resemble each other more closely, or respond more similarly to the study intervention, than persons drawn from different clusters. It should also be checked whether the randomization for the study has been carried out with such methods as stratification and covariate-adjusted randomization. CRT can be analyzed on either the individual or the cluster level. The rationale for the choice of a cluster-randomized design should be explained, and intracluster correlation coefficients (ICC) should be reported as an aid to the planning of future studies. Particular requirements are also described in an extended version of the CONSORT guidelines that has been developed specifically for CRT.

Conclusion: Readers of publications on CRT should be aware of the special requirements mentioned above with respect to the design, performance, and analysis of this type of study as opposed to individually randomized studies. If no special techniques are applied in the design, performance, and analysis of a CRT, or if the assumptions underlying each of these steps have not been properly checked, then the findings of the study may well be misleading.

Cite this as:

Lorenz E, Köpke S, Pfaff H, Blettner M: Cluster-randomized studies—part 25 of a series on evaluating scientific publications. *Dtsch Arztebl Int* 2018; 115: 163–8. DOI: 10.3238/arztebl.2018.0163

Cluster-randomized trials (CRT) are often carried out to evaluate the kind of complex interventions that are increasingly being adopted in health services research, for example (1). Complex interventions consist of several individual interventions that may interact with each other. An example from Germany is a study on guideline-based reduction of the use of physical restraints (PR) in residential care homes, where information sheets were distributed, training courses given, and PR officers designated in the facilities involved (*Box 1*).

The number of publications concerning themselves with CRT has increased continuously over the past 10 years (the PubMed search term “cluster randomised trial” OR “cluster-randomised trial” OR “cluster randomized trial” OR “cluster-randomized trial” threw up 54 hits in 2006, 156 in 2011, and 392 in 2016), and articles on CRT formed a four times higher proportion of all Medline-indexed studies in 2016 than they did in 2006.

In a CRT, not individual participants but rather whole facilities or groups of participants (clusters) are allocated to one or more intervention or control group (2). For instance, in the above-mentioned study on reduction of PR, all study participants under the care of a given physician received the same information and the same ensuing intervention, without being influenced by other participants who received the control intervention (3). Formation of clusters was necessary in this study (*Box 1*) because the residents of a care home cannot be considered as independent of each other.

A study may have to be conducted as a CRT if an intervention is being performed not at individual level but at the level of whole regions or organizations. Examples of such situations are the restructuring of a hospital, the implementation of guidelines, or the introduction of a new form of care. The intervention cannot be withheld from any individual in the organizational unit for fear of contamination. In the case of the intervention to reduce physical restraints in residential care homes, individual randomization is impossible because intervention addresses the care given in the whole facility.

This article describes design considerations, randomization strategies, statistical methods, and the strengths and limitations of CRT. Our aim is to enable the reader to scrutinize and interpret the results of CRT in critical fashion, taking into account the methodological requirements.

Institute for Medical Biostatistics, Epidemiology and Informatics, Mainz University Medical Center: Dr. Lorenz, Prof. Blettner

Department of Teaching and Research in the Care Sector, Institute for Social Medicine and Epidemiology, University of Lübeck: Prof. Köpke

Institute for Medical Sociology, Health Services Research, and Rehabilitation Science, University of Cologne: Prof. Pfaff

Center for Health Services Research Cologne (ZVFK), University of Cologne: Prof. Pfaff

BOX 1

Cluster-randomized trials (CRT) in residential care homes for the elderly

Despite legal restrictions and evidence of deficiencies in effectiveness and safety, physical restraints (PR) are often employed in care homes. In the framework of a CRT, a guideline-based complex intervention was carried out to reduce the use of PR, including distribution of information materials, training courses, and appointment of PR officers in each facility. The intervention addressed processes in the whole facility. Randomization at the level of individual residents is therefore not feasible owing to the danger of contamination. The primary endpoint was the prevalence of PR 6 months after commencement of the intervention, as determined in the course of direct, unannounced visits by blinded external study assistants.

Sample calculation: For reduction of PR prevalence, an change from 33% to 21% in 6 months was assumed, with power of 90% at a significance level of 5% in a two-sided test. On the basis of earlier studies the intracluster correlation coefficient (ICC) was estimated at 0.034, resulting in a design factor of 5.0 and thus a necessary sample size of 2824 residents in 34 care homes. The mean cluster size was 83 residents. In anticipation of a drop-out rate of 5% of facilities and 2% of residents (not including deaths and moves away), the final sample was fixed at 3060 residents in 36 care homes.

Avoidance of bias: Prospective publication of the study protocol, stratified concealed block randomization after acquisition of the baseline data of all participating residents, inclusion of all persons resident in each facility at the time of the survey by waiving individual consent, intensified (and successful) efforts to avoid loss of clusters, blinded documentation of the primary endpoint by external study assistants including verification of successful blinding.

Results: All care homes completed the study and all residents were included in the analyses. The prevalence of PR went down from 30.6% to 29.1% in the control group and from 31.5% to 22.6% in the intervention group over the 6-month study period.

Conclusion: The guideline-based complex intervention reduced the use of PR in care homes.

(Modified from Köpke et al. 2012 [18])

Planning

Power and case number calculation

The formation of clusters decreases the effective sample size and thus the statistical power of CRT compared with individually randomized trials, because persons within an organizational unit resemble each other more strongly than persons from different organizational units. Similarity within and between clusters is quantified with the aid of the intracluster correlation coefficient (see equation in *Box 2*). For the purposes of calculation the clusters are assumed to be the same size; extension to take account of unequal cluster sizes is possible (4).

The resulting necessity to enlarge the sample size is known as the design effect (DE) (see equation in *Box 2*) (5). Thus, a DE of 1.2 means that the cluster size, assuming equally large clusters, has to be increased by 20% compared with individual randomization. If there is no correlation within the cluster (i.e., intracluster correlation coefficient [ICC] = 0), then DE = 1 and the sample size of the CRT corresponds to that of a study with individual randomization. Conversely, if all elements in a cluster react to the intervention in the same way (i.e., ICC = 1), then the number of clusters needed is the same as the number of individuals in an individually randomized study (*Table*).

- If the DE is ignored when planning a CRT, type 2 error increases (*Box 3*).
- Ignoring the DE and the resulting decrease in variance in the course of analysis at cluster level leads to an increase in type 1 error (*Box 3*).

Case number planning necessarily includes deciding on the number of clusters and the number of individuals per cluster. The latter is often determined by the basic investigation unit and may vary widely (e.g., number of residents in a care home for the elderly). In addition to the usual assumptions in case number calculation, an assumption of the anticipated ICC is needed. The required sample size increases with increasing ICC.

One must also take into consideration that the power can no longer be increased very much when the number of individuals per cluster exceeds 1/ICC (5). In other words, it is not helpful to include a lot of “large” clusters in one’s study. In this case a random sample can be selected within a cluster, meaning that fewer individuals are included than are available in that cluster. With an ICC of 0.05, a total of around 20 (1/0.05) persons per cluster suffices to achieve the desired power. There are often no valid data on the ICC, in which case a literature-based figure or a realistic value based on earlier studies should be used. Experience shows that an ICC of around 0.05 can be assumed for studies of primary care (6); in community-randomized studies the ICC is usually lower (0.01 or often even 0.001 [2]). The resulting DE is nevertheless large in large clusters.

Furthermore, the necessary number of cases depends on the size of the clusters: 100 clusters each containing 10 probands lead to greater statistical power than 10 clusters of 100 probands each. Recruitment of additional clusters yields greater effective case numbers than recruitment of more individuals in

BOX 2

Equations

● Intracluster correlation coefficient (ICC)

The ICC is defined as

$$ICC = \frac{s_b^2}{(s_b^2 + s_w^2)}$$

where s_b^2 is the variability among clusters and s_w^2 is the variability within clusters.

● Design effect (DE)

The DE is defined as

$$DE = 1 + (\bar{m} - 1) \times ICC$$

where \bar{m} is the average cluster size.

● Randomization scheme

The number of possible randomization schemes is given as $\binom{n}{k}$, where n is the total number of clusters and k the number of clusters per study arm.

TABLE

Effective sample size and power at constant total sample size with different numbers of clusters, numbers of patients per cluster, intracluster correlation coefficients, and design effects

Clusters k	Patients m	Total (mk)	ICC	DE	ESS	Power t-test*
144	1	144	0.00	1.00	144	85
72	2	144	0.01	1.01	142	84
72	2	144	0.05	1.05	137	82
36	4	144	0.01	1.03	140	82
36	4	144	0.05	1.15	125	78
18	8	144	0.01	1.07	135	78
18	8	144	0.05	1.35	107	68
8	18	144	0.01	1.17	123	67
8	18	144	0.05	1.85	78	48

* Power in % for demonstration of an effect size of 0.5 between clusters assigned equally to intervention and control (two-sided t-test in studies with two arms of the same size; $\alpha = 0.05$; standard deviation SD = 1). DE, Design effect; ESS, effective sample size; ICC, intracluster correlation coefficient; k, number of clusters; m, number of patients per cluster; mk, total number of patients included

clusters (7). As an ad-hoc approach, the number of cases for individual randomized trials can be calculated and multiplied by the DE. The formula has to be expanded in the case of extreme variation in cluster size (8).

A large number of strategies for case number planning in CRT have been published, and implemented with the aid of various statistical analysis softwares such as R and Stata (9–13).

Potential pitfalls in planning and sources of bias in cluster-randomized trials

The Cochrane Handbook (Higgins & Green 2011, [14]) lists four specific potential sources of distortion in the context of CRT:

- Recruitment bias
- Baseline imbalance among groups
- Loss of clusters
- Incorrect analysis

Distortion can arise as early as the recruitment stage, if participants cannot be followed-up for the whole duration of the study or the intention-to-treat (ITT) analysis is not carried out. To avoid this source of bias, it should be ensured that data can be acquired from all members of the randomized clusters (or of the random sample). In the event of incomplete follow-up, techniques for dealing with missing data should be employed (15).

Allocation concealment (blinding; see Box 3) is often not feasible in a CRT, where bias can arise through intervention assignment. For instance, the motivation of study staff to recruit patients can de-

pend on the intervention arm. Equally, the patients' motivation to take part may be affected by previous knowledge of the various interventions that are to be compared. Brierley et al. published a review of susceptibility to recruitment bias (16). To avoid this source of distortion, the recruitment of study participants should be completed before randomization. Because the study staff and patients often cannot be blinded, at least the documentation of the primary outcome parameter should be accomplished by others.

Trial conduct

Randomization

Equal distribution of potential influencing factors and sources of disturbance is a precondition for being able to attribute observed effects to an intervention. The units of randomization in a CRT can be care homes (see example), hospital groups, hospitals, hospital wards, doctors' offices, schools, or whole local communities. These groups do not arise by chance but as the result of social, geographic, or other interacting factors. Various randomizing strategies exist to nevertheless ensure even distribution.

Simple randomization

In simple (unrestricted) randomization, the clusters are assigned randomly to the treatment and control arms. In the case of a small number of clusters of varying size, this may result in wide discrepancies in sample size. Considerable imbalance of study participants' characteristics can arise both at cluster level and at the individual level.

BOX 3

Definitions of central terms

- Type 1 error: A type 1 error is present when the null hypothesis is wrongly rejected.
- Type 2 error: A type 2 error is present when the null hypothesis is wrongly retained.
- Blinding: This means that study participants, study physicians, other study staff, or assessors are not aware of the assigned intervention and are therefore not influenced by that knowledge.
- Allocation concealment: With concealed allocation, the persons who carry out randomization do not know to which study group the next participant will be allotted.

Matching

To prevent blatant mismatching of the clusters in the intervention and control groups from the outset, the participating clusters are paired with regard to factors such as age, sex, cultural background, socioeconomic status, and occupation. In our example, randomization would be preceded by formation of “cluster pairs,” each comprising two care homes with similar age structure and sex distribution (*Figure*). In each cluster pair, one cluster is randomly selected for the intervention, thus guaranteeing that the two arms of the trial are balanced. However, this means that whenever a cluster leaves the study (loss to follow-up), the paired cluster has to be excluded. To alleviate this problem, matching can be set aside at the data analysis stage (17).

Stratification

In the case of stratification the study population is divided into disjoint groups (“strata”). In the study by Köpke et al., study regions 1 and 2 were stratified for randomization (*Figure*) (18). Each stratum is homogeneous with regard to relevant characteristics, but the strata may differ very widely from one another. Clusters for the intervention and control arms are chosen randomly to form equally sized blocks in each stratum. The number of strata should be kept low so that balanced blocks result. This requirement often stands in opposition to the frequent need for randomization to take account of a large number of variables by differentiation of cluster and individual level. For example, stratification in a geographic region with four values and two funding bodies would involve division of the clusters into eight strata. Such a strategy can lead to underoccupation of individual cells.

Minimization

The minimization method represents a compromise between balance and (true) randomization. Individual clusters are allocated sequentially to the intervention arm and the control arm while taking account of the

participants’ relevant characteristics. The aim is to make the arms of the trial as homogeneous as possible. In the case of a small number of clusters this balance runs against the principle of randomness and may lead to an increased risk of selection bias. In minimization the number of covariables for stratification is limited, so that the variables that are considered can also be modeled when it comes to analysis. Clusters are deterministically assigned to an intervention or the control group according to relevant variables. In this way observable confounders can be balanced between the study arms.

Covariable randomization

Another approach is covariable-restricted randomization, in which clusters are allotted to the study arms in equal numbers according to the distribution of relevant basic variables (19–21). For constant variables one takes account of aggregated data such as mean values within clusters or strata. Data from the basic data acquisition stage must already be available at the time of randomization. A randomization scheme is selected randomly from among those that result in balanced study arms with regard to predefined relevant properties and exposures. Because the final randomization scheme is selected from the group of all theoretically possible schemes (see the equation for the number of possible randomization schemes in *Box 2*), randomness of assignment to intervention or control is largely preserved.

The evaluation of CRT takes place on at least two levels, namely the cluster level and the individual (patient) level. In the multi-level models the statistical model is expanded by adding a random component for the variation of the clusters (21, 22). This takes account of the ICC resulting from the design. A lucid account of how to carry out a multi-level analysis is provided by Ansmann et al. (23).

Presentation

The use of CRT has increased steeply in the past 15 years. This has led to expansion of the CONSORT guidelines (www.consort-statement.org/) on the publication of trials of this type, because CRT design presents specific methodological challenges (24, 25). The principal extensions of the CONSORT guidelines with regard to CRT are as follows:

- The reasons for deciding to perform a CRT should be explicitly laid out.
- The provision for the influence of clustering in the individual phases of the study from case number calculation through randomization to analysis should be described.
- The ICC should be presented as a basis for case number planning in future studies.

The reporting of CRT in medical research currently displays major deficits. Therefore, it is very important that authors plan their studies in accordance with the expanded CONSORT guidelines or, for example, the stepped wedge design (26).

Discussion

The first step in planning a study is to decide whether it can be performed with individual randomization or whether a CRT is necessary. An acceptable reason for carrying out a CRT is that the intervention is being performed in clusters and there would be a risk of contamination in an individually randomized trial. Alternative study types for organizational interventions are the stepped wedge and crossover designs, in which the clusters are included in analysis both as intervention and as control entities.

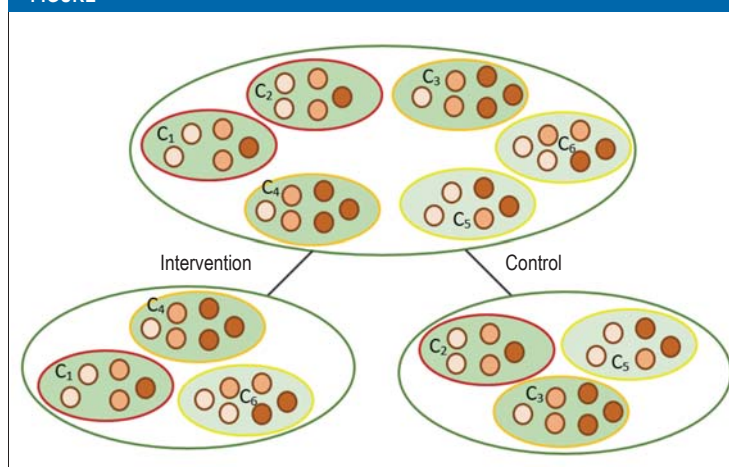
The planning and conduct of CRT presents special challenges differing from the requirements for individually randomized trials. The clustering must be retained at all stages, from case number planning through analysis techniques to reporting. Study conduct also involves specific challenges, e.g., regarding selection bias and information bias.

Whether conclusions should be drawn at the individual patient level or at cluster level is determined by the choice of design and analysis technique (5). To increase analytical precision, strict inclusion and exclusion criteria have to be defined. This can be achieved, for example, by recruiting physicians' offices of similar size or physicians with similar professional experience. It is always important to consider the benefit of an intervention not only at cluster level but also at the level of individual patients, e.g., improvement in quality of life by reduction of PR or improvement in the reputation of the care home with fewer complaints from relatives.

When planning a study, the study-specific ICC can be estimated on the basis of a baseline survey. Moreover, stratification variables that are already relevant can be identified. Because CRT are frequently used to evaluate "complex" interventions (27), one should adhere to the corresponding guidelines, such as those of the UK Medical Research Council (MRC) (28).

It is often argued that the conduct of CRT is associated with less administrative effort, e.g., in connection with the acquisition of aggregated data. On the other hand, the consent of the study participants must be obtained at two levels, because although the intervention is carried out at cluster level, when it comes to analysis there are frequently interesting parameters at the individual level. For large communities, it may be logistically challenging or even impossible to obtain informed consent for all individual study participants (5, 29). However, this should not necessarily be viewed as a limitation of ethical requirements, provided there is sufficient justification (30). The level at which consent is necessary depends on the intervention, on the study-specific data protection regulations, and on the specific requirements of the ethics committee involved. In some situations it may be justified to go ahead in the absence of informed consent, for instance if the intervention only tangentially affects individual persons. An example is the introduction of new rules regarding hygiene, which does not require the agreement of all patients.

FIGURE



Matching and stratification: The clusters shaded dark green (region 1) and light green (region 2) represent the two study regions in our example of a study of care homes. The different-colored contours (red, orange, yellow) of the clusters show homogeneous cluster pairs. The various colors of the individuals in each cluster (light red, medium red, dark red) identify the various age groups.

Matching in cluster-randomized studies: In the matching process, the maximally homogeneous cluster pairs are divided randomly between the intervention and control arms of the study with regard to the predefined participant characteristics (in this example: age).

Stratification in cluster-randomized studies: In stratified randomization, care homes from a region are randomly selected from the whole set of care homes in that region and divided equally between the two study arms.

Acknowledgments

We are grateful to Prof. Lena Ansmann and Michael Swora for their valuable comments which helped to improve the quality of this article.

Conflict of interest statement

The authors declare that no conflict of interest exists.

Manuscript submitted on 12 June 2017, revised version accepted on 26 October 2017

Translated from the original German by David Roseveare

References

- Hayes RJ, Moulton LH: Cluster randomized trials. 1 ed: Chapman and Hall/CRC Press, Boca Raton, FL; 2009.
- Donner A, Klar N: Design and analysis of cluster randomization trials in health research. London: Arnold Publishers Limited 2000.
- Bland JM, Kerry SM: Trials randomised in clusters. *BMJ* 1997; 315: 600.
- Eldridge SM, Ashby D, Kerry S: Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol* 2006; 35: 1292–300.
- Donner A, Klar N: Pitfalls of and controversies in cluster randomization trials. *Am J Public Health* 2004; 94: 416–22.
- Campbell MJ: Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000; 9: 81–94.
- Flynn TN, Whitley E, Peters TJ: Recruitment strategies in a cluster randomized trial—cost implications. *Stat Med* 2002; 21: 397–405.
- Kerry SM, Bland JM: Unequal cluster sizes for trials in English and Welsh general practice: implications for sample size calculations. *Stat Med* 2001; 20: 377–90.
- Campbell MK, Thomson S, Ramsay CR, MacLennan GS, Grimshaw JM: Sample size calculator for cluster randomized trials. *Comput Biol Med* 2004; 34: 113–25.
- Donner A, Birkett N, Buck C: Randomization by cluster. Sample size requirements and analysis. *Am J Epidemiol* 1981; 114: 906–14.

Key messages

- Cluster-randomized trials (CRT) are frequently used when interventions are to be carried out at the level of whole groups rather than single individuals.
- As a result of cluster formation, persons within a group often have more characteristics in common than persons in different groups. The so-called intracluster correlation coefficient should be reported as measure of similarity of individuals between and within clusters.
- The results of CRT can be distorted by recruitment bias, baseline imbalance, loss of clusters, and incorrect analysis.
- Blinding of the participants and study personnel in a CRT is frequently not feasible. This may result in differing motivation and thus become a source of recruitment bias. If at all possible, therefore, recruitment of study participants should be complete before randomization.
- In CRT where the study staff are not blinded, the outcome parameters should be documented by an external person.

11. Feng Z, Diehr P, Peterson A, McLerran D: Selected statistical issues in group randomized trials. *Annu Rev Public Health* 2001; 22: 167–87.
12. Guittet L, Giraudeau B, Ravaud P: A priori postulated and real power in cluster randomized trials: mind the gap. *BMC Med Res Methodol* 2005; 5: 25.
13. Kerry SM, Bland JM: Sample size in cluster randomisation. *BMJ* 1998; 316: 549.
14. Higgins J, Green S (eds.): *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011. <http://handbook.cochrane.org> (last accessed on 1 December 2017).
15. Giraudeau B, Ravaud P: Preventing bias in cluster randomised trials. *PLoS Medicine* 2009; 6: e1000065.
16. Brierley G, Brabyn S, Torgerson D, Watson J: Bias in recruitment to cluster randomized trials: a review of recent publications. *J Eval Clin Pract* 2012; 18: 878–86.
17. Diehr P, Martin DC, Koepsell T, Cheadle A: Breaking the matches in a paired t-test for community interventions when the number of pairs is small. *Stat Med* 1995; 14: 1491–504.
18. Köpke S, Mühlhauser I, Gerlach A, et al.: Effect of a guideline-based multicomponent intervention on use of physical restraints in nursing homes: a randomized controlled trial. *JAMA* 2012; 307: 2177–84.
19. Ivers N, Halperin I, Barnsley J, et al.: Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. *Trials* 2012; 13: 120.
20. Lorenz E, Gabrysch S: ccrand: Covariate-constrained randomization routine for achieving baseline balance in cluster-randomized trials. *Stata J* 2017; 17: 503–10.
21. Moulton LH: Covariate-based constrained randomization of group-randomized trials. *Clin Trials (London, England)* 2004; 1: 297–305.
22. Diez-Roux AV: Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998; 88: 216–22.
23. Ansmann L, Kuhr K, Kowalski C, für die Arbeitsgruppe Organisationsbezogene Versorgungsforschung des DNVF: Mehrebenenanalysen in der organisationsbezogenen Versorgungsforschung – Nutzen, Voraussetzungen und Durchführung. *Gesundheitswesen* 2017; 79: 203–9.
24. Campbell MK, Elbourne DR, Altman DG: CONSORT statement: extension to cluster randomised trials. *BMJ* 2004; 328: 702–8.
25. Campbell MK, Piaggio G, Elbourne DR, Altman DG: Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; 345: e5661.
26. Bland JM: Cluster randomised trials in the medical literature: two bibliometric surveys. *BMC Med Res Methodol* 2004; 4: 21.
27. Mühlhauser I, Lenz M, Meyer G: Bewertung von komplexen Interventionen: Eine methodische Herausforderung. *Dtsch Arztebl* 2012; 109: A 22–3.
28. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M: Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008; 337: a1655.
29. Giraudeau B, Caille A, Le Gouge A, Ravaud P: Participant informed consent in cluster randomized trials: review. *PLoS One* 2012; 7: e40436.
30. Sim J, Dawson A: Informed consent and cluster-randomized trials. *Am J Public Health* 2012; 102: 480–5.

Corresponding author

Dr. sc. hum. Eva Lorenz
 Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI)
 Universitätsmedizin der Johannes Gutenberg Universität Mainz
 Obere Zahlbacher Str. 69
 55131 Mainz, Germany
 eva.lorenz@uni-mainz.de