BMJ Open

How often are interventions in cluster-randomised controlled trials of complex interventions in general practices effective and reasons for potential shortcomings? Protocol and results of a feasibility project for a systematic review

Andrea Siebenhofer,[1,2] Stefanie Erckenbrecht,[3] Gudrun Pregartner,[4] Andrea Berghold,[4] Christiane Muth[1]

To cite: Siebenhofer A, Erckenbrecht S, Pregartner G, *et al*. How often are interventions in cluster-randomised controlled trials of complex interventions in general practices effective and reasons for potential shortcomings? Protocol and results of a feasibility project for a systematic review. *BMJ Open* 2016;**6**:e009414. doi:10.1136/bmjopen-2015-009414

▸ Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2015-009414).

AS and SE contributed equally.

Received 16 July 2015
Revised 11 September 2015
Accepted 23 October 2015

For numbered affiliations see end of article.

Correspondence to
Professor Andrea Siebenhofer; andrea.siebenhofer@medunigraz.at

ABSTRACT

Introduction: Most studies conducted at general practices investigate complex interventions and increasingly use cluster-randomised controlled trail (c-RCT) designs to do so. Our primary objective is to evaluate how frequently complex interventions are shown to be more, equally or less effective than routine care in c-RCTs with a superior design. The secondary aim is to discover whether the quality of a c-RCT determines the likelihood of the complex intervention being effective.

Methods and analysis: All c-RCTs of any design that have a patient-relevant primary outcome and with a duration of at least 1 year will be included. The search will be performed in three electronic databases (MEDLINE, EMBASE and the Cochrane Database of Systematic Reviews (CDSR)). The screening process, data collection, quality assessment and statistical data analyses (if suitably similar and of adequate quality) will be performed in accordance with requirements of the Cochrane Handbook for Systematic Reviews of Interventions. A feasibility project was carried out that was restricted to a search in MEDLINE and the CCTR for c-RCTs published in 1 of the 8 journals that are most relevant to general practice. The process from trial selection to data collection, assessment and results presentation was piloted. Of the 512 abstracts identified during the feasibility search, 21 studies examined complex interventions in a general practice setting. Extrapolating the preliminary search to include all relevant c-RCTs in three databases, about 5000 abstracts and 150 primary studies are expected to be identified in the main study. 14 studies included in the feasibility project (67%) did not show a positive effect on a primary patient-relevant end point.

Ethics and dissemination: Ethical approval is not being sought for this review. Findings will be disseminated via peer-reviewed journals that frequently publish articles on the results of c-RCTs and through presentations at international conferences.

Trial registration number: PROSPERO CRD201400923.

Strengths and limitations of this study

- Study selection, data extraction, and the assessment of risk of bias will be conducted by two authors independently.
- A comprehensive feasibility check was carried out for a systematic review and the results provided important information on how to design the study.
- It will be difficult to pool data because the target population is very variable and not limited to specific conditions or diseases. There is also likely to be considerable variation in patient-relevant primary outcomes.
- No search in trial registries to indicate possible publication bias is planned.

INTRODUCTION

Primary care delivered at general practices is critical in any healthcare system, and its importance is increasing due to the rising prevalence of chronic diseases and multimorbidity in an ageing population.[1] At the same time, the need to control healthcare costs makes it particularly important that healthcare professionals use interventions with proven effectiveness.

Most studies conducted at general practices focus on the behaviour of patients and health professionals, or on organisational

change[2] and concentrate on interventions such as disease management programmes, vaccination programmes, and screenings. As such interventions are generally complex, cluster-randomised controlled trials (c-RCTs) are increasingly used to evaluate them. However, c-RCTs have certain methodological shortcomings, a common example of which is inadequate concealment of the treatment allocation. As a result, the CONSORT statement was updated and extended to c-RCTs in 2004 and 2010, and now includes specific advice on how to meet various quality standards.[3 4] In addition, the 2013 Ottawa Statement describes the ethical issues that should be considered when conducting c-RCTs and provides guidance and key recommendations for researchers and ethics committees.[5]

The present manuscript describes the protocol of a methodological systematic review on the basis of a feasibility project. It has the primary objective of evaluating how frequently complex interventions are shown to be more, equally or less effective than routine care in c-RCTs that use a superior design. The secondary objective of the review is to discover whether the quality of a c-RCT determines the likelihood that the complex intervention will be proven to be effective.

METHODS

Criteria for inclusion in this review

Eligibility criteria

All c-RCTs involving adults, adolescents, and children in a general practice setting will be included. The trials must investigate a complex intervention in accordance with the recommendations of the latest Medical Research Council (MRC) guidance: we have included all interventions that involve interacting components in the experimental group—such as treatment, changes in behaviour required by those delivering and/or receiving interventions, and/or changes in the number of organisational levels targeted by the intervention.[6] To avoid additional heterogeneity between studies arising from active comparators, the control group must have continued to receive treatment as usual (routine care). For inclusion in our review, trialists must either have explicitly defined primary outcome(s) as primary or main outcome(s), have used such outcome(s) in a power and sample size calculation, or have listed it (these) as the main outcome(s) in their trial's objectives.[7] In addition, the primary outcome(s) has (have) to be patient relevant, and detailed criteria for the assessment of the patient-relevant end points should have been determined in accordance with the Institute for Quality and Efficiency in Health Care (IQWiG) methods V.4.2, which give a concise and literature-based definition of what is meant by patient relevance.[8] In this connection, 'patient relevant' refers to how a patient feels, functions, or survives—that is, whether indicators of mortality, morbidity, health-related quality of life, hospitalisation and/or treatment satisfaction are provided. If a study reports on more than one primary end point, only the patient-relevant end points will be included in this study. As we want to evaluate the long-term benefit of an intervention, only studies of at least 12 months' duration will be considered. Inclusion and exclusion criteria are shown in table 1.

Outcome measures

1. Summarising the evidence from c-RCTs to describe the distribution of estimates of treatment effect with respect to direction (in favour of the complex intervention or the routine care treatments), magnitude (size of the effect), and statistical significance (or CI).
2. Evaluating how frequently complex interventions in c-RCTs are more, equally or less effective than routine care.
3. Exploring the extent to which methodological (eg, power calculations and intra-cluster correlation coefficients (ICC)) and other factors (eg, ethical approval, sponsorship, run-in phase) explain differences in the

Table 1 Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
All c-RCTs of any design (eg, parallel, cross-over, or stepped wedge)	Main results were previously published; design papers; protocols or pilot studies and 'side papers' such as tertiary literature
Superiority trials	There is no reporting on patient-relevant primary outcomes
	Only practice (general practice level) data are available
c-RCT compares a complex intervention with routine care	No consideration of comparators, such as active controls or sham interventions, will be permitted due to homogeneity of samples
Studies have a patient-relevant primary outcome (no surrogates)	
Study duration >12 months	
Study included all age groups	
Studies examine individual patient data	
Funding is not relevant (commercial, non-commercial, other funding)	
General practice is the cluster	
No language restrictions	
Published studies only	

c-RCT, cluster-randomised controlled trail.

distribution of c-RCTs that show results that either favour or disadvantage complex interventions.

Search methods

The search strategy was developed by the Institute of General Practice at Goethe University Frankfurt, Germany, in cooperation with the Centre for Research in Evidence-Based Practice at Bond University, Australia, and was broadly based on a validated approach developed by Taljaard *et al*.[9] Relevant papers will be identified by searching the Central Register of Controlled Trials (CCTR) (last issue), MEDLINE (from 1962 until recently) and EMBASE (from 1988 until recently) without any language restriction. The full search strategy for MEDLINE and CCTR appears in table 2 and will be adapted for EMBASE. The proposed end date for the literature search is September 2015.

The literature search for the feasibility project was carried out in the databases 'Cochrane Central Register of Controlled Trials' (EBM Reviews—CCTR) and 'MEDLINE' between 1946 and April 2014. Following recommendations by Eldridge *et al*,[10] the search strategy for the feasibility project was restricted to the journals publishing the highest numbers of articles related to general practice, namely the *British Medical Journal, British Journal of General Practice, Family Practice, Preventive Medicine, Annals of Internal Medicine, Journal of General Internal Medicine and Paediatrics*. The *Canadian Medical Journal (CMJ)* was also included because the initial inspection of a 10% sample of the unrestricted search results showed that this journal also contains a high number of c-RCTs in a general practice setting. For the full review, the search will not be restricted to articles published in 'general practice'-related journals.

To ensure literature saturation, we will scan references in methodological and relevant secondary literature that were identified in the three electronic databases and reference lists of the included studies and that were published after January 2010. We will also search the authors' personal files—literature collected during the conceptual development of our project idea—to make sure that all relevant material has been captured.

Table 2 Search strategy for MEDLINE and CCTR

	Databases: MEDLINE (Ovid) (1946 to April 2014) and EBM Reviews—Cochrane Central Register of Controlled Trials (Ovid) (January 2014)
1	general practitioners/ or physicians, family/ or physicians, primary care/
2	Primary Health Care/
3	exp General Practice/
4	((family or general or primary) adj3 (practic* or practition*)).tw.
5	primary care.tw.
6	(gp or gps).tw.
7	((family or primary or general) adj3 (physician* or doctor* or clinician*)).tw.
8	or/1–7
9	(cluster* adj2 randomi*).tw.
10	exp cluster analysis/
11	(practice* adj5 random*).tw.
12	or/9–11
13	exp animals/ not humans.sh.
14	12 not 13
15	8 and 14
16	remove duplicates from 15

Expected primary studies

The initial search of the feasibility project identified 512 papers. Of these, 426 were excluded following abstract screening. The full texts of the remaining 86 studies were screened and 21 papers, or 4% of initial findings, ultimately fulfilled the inclusion criteria (see online supplementary file 1). When these results are extrapolated to take account of an unrestricted literature search that includes EMBASE as a third database as well as journals that do not focus on general practice, we expect to have about 5000 findings and 150 papers (3% of initial findings).

Expected authors' responses

Eighteen authors were contacted for further information on the intra-cluster correlation coefficients (ICCs) assumed in the sample size calculation and observed in the data used in their studies; three authors had already provided the necessary information in their publications. If no response was forthcoming, reminders were sent 4 weeks after the initial contact. Of the 18 authors we contacted, 4 forwarded the relevant information; thus, ICCs were available for about 22% of studies (see online supplementary file 2).

Selection of trials and data collection

The abstracts, titles and full texts will be independently screened by two reviewers. Data from each study will be assessed independently by the two authors and entered into data extraction templates. Disagreements will be resolved by a third reviewer and relevant missing information will be requested from the original authors of the study.

Quality assessment

The criteria listed were developed during the feasibility phase and based on the CONSORT statement—extension to cluster randomised trials;[4] the extraction sheets for RCTs used by IQWiG;[11] the Cochrane Handbook for Systematic Reviews of Interventions;[12] and the systematic review of Froud *et al*.[13] The authors extracted all named criteria and then decided which to include in the assessment, based on their frequency and relevance to the research question. Additional criteria, such as number of participating practices and length of observation period for intervention and control groups, were included because these were considered to be important

quality measures. Finally, the reported criteria were grouped thematically (general information, sample size calculation, randomisation and blinding process, analyses) (see online supplementary file 3) and piloted using the studies identified during the feasibility phase. The results of the preliminary assessment are presented in online supplementary file 3 which includes tables representing the final template for the upcoming full review. As additional information, we will extract from the discussion section of the included studies the authors' own interpretations and explanations as to why their studies did not show a positive effect.

Data analysis

Data will be summarised (or pooled) statistically where appropriate. We will perform the statistical analyses in accordance with the guidelines provided in the latest version of the Cochrane Handbook for Systematic Reviews of Interventions.[12] In addition, descriptive plots and analyses will be performed to explore the distribution of effect sizes, and the frequency of complex interventions in c-RCTs being more, equally or less effective than routine care. Data analysis will be performed in Cochrane Review Manager 5.1.0. We will use either HR or OR to estimate the individual and overall effects of studies that are presented with a 95% CI. We will also calculate the heterogeneity statistics (χ^2 and I^2), and test the robustness of the results by repeating the analysis using different statistical models (fixed-effect and random effects model). When heterogeneity is found, we will attempt to determine the reasons for this by examining individual study and subgroup characteristics. Subgroup analyses are planned if sufficient RCTs can be identified, for example, on study fields, type of outcome, and type of practice. We will perform sensitivity analyses in order to explore the robustness of our results and visually inspect funnel plots for any indication of publication bias.

The intra-cluster correlation coefficient (ICC) gives a measure of the similarity of observations from the same class. It is usually defined as the proportion of variance accounted for by class variation and can be estimated by analysing variance methods.[14] To assess whether unrealistic assumptions regarding the ICC used in the sample size calculation may have resulted in trials failing to show the superiority of a complex intervention, we will compare the available ICC pairs used for the sample size calculation with the ICCs actually obtained from the data. Furthermore, descriptive statistics such as minimum, maximum and median absolute differences will be stated.

ETHICS AND DISSEMINATION

Ethics approval will not be required, since this is a protocol for a systematic review utilising published data. Results will provide information on the shortcomings of c-RCTs and help in the design of studies with complex interventions. Once completed, the results from this systematic review will be published in a peer-reviewed journal, and presented at international and national conferences.

DISCUSSION

During the course of the development of the final protocol, the feasibility project was extraordinarily useful for judging whether the planned systematic review would be feasible in terms of the numbers of trials expected, the definition of the inclusion and exclusion criteria, the development of data extraction forms, and results presentations. It also helped summarise the quality criteria that need to be collected to deal with the question of which methodological and other factors explain differences in c-RCTs to show results that either favour or disadvantage complex interventions. Using a selective search in journals relevant to general practice, a total of 21 studies examining complex interventions in a general practice setting were identified during the feasibility project. Fourteen of these did not demonstrate a positive effect on a primary patient-relevant end point. This corresponds to 67% of all studies, which considering that a complex intervention study usually requires considerable effort and monetary resources—as well as a large number of patients—is a striking number.

In order to describe and analyse the differences between studies with and without an intervention effect, we developed our own checklist which includes 18 quality aspects based on previously used criteria in other methodological papers. For example, the CONSORT Statement requires that both estimated and observed ICCs are reported, and it is interesting to note that several papers providing lists of ICCs that are relevant to general practice have been published and recommended for use in study design.[15 16] However, in our feasibility project we only found three studies[17–19] that quoted both ICC values, indicating that journals and reviewers do not attach enough importance to this issue and should request this information more rigorously. When the search is expanded to include three databases and with no restriction to general practice-related journals, we expect this methodological review to provide an answer to the research question. A minor limitation for this study that should be noted is that no search in trial registries is planned, and the existence of publication bias can, therefore, not be ruled out. We selected complex interventions as the vast majority of interventions at general practices are multifaceted. The term 'complex intervention' is thus a descriptive element of the study. In a subsequent substudy—based on samples of the included studies—we will appraise the complex interventions themselves in accordance with the recommendations of Möhler *et al*.[20 21]

Diaz-Ordaz[22] published a review based on 73 c-RCTs conducted in residential facilities for older people. Less than 30% of the trials had accounted for clustering in their sample size calculations, and considerable

differences existed between studies with and without an intervention effect. Another recently published systematic review by Ivers *et al*[23]—that dealt with more than 300 cluster-randomised trials published between 2000 and 2008—clearly showed that despite the publication of the CONSORT statement on the reporting and methodological quality of c-RCTs in 2004,[3] very few aspects had been adhered to and further effort was necessary to improve methodological quality. A Cochrane Review, which was not restricted to c-RCTs and was published by Turner in 2013,[24] demonstrated that in RCTs, the factors named in the CONSORT statement were more often fully reported when journals had actively encouraged its use.

Several further reviews published before 2005 show that even though there has been some methodological improvement in terms of appropriate sample size calculations and analyses, weaknesses are still present, especially with regard to blinding and allocation status.[2 10 25] In addition to these methodological reviews, the publication of the extended CONSORT statement[3] and the Ottawa Statement[5] also help researchers to avoid pitfalls and design c-RCTs properly.

To the best of our knowledge, our research question—which aims to evaluate whether, and if so under what circumstances, it is sensible to assess the efficacy of a proposed intervention by conducting a c-RCT in a general practice setting—is novel. Our feasibility project has enabled us to obtain a valid estimate of the proportion of studies that are effective, and this provides the basis for a project that has been approved by the German Federal Ministry of Education and Research, and registered with Prospero (CRD42014009234). By performing a full literature search and exploring the extent to which methodological (eg, power calculation, intra-cluster correlation, handling of missing data[25]) and other factors (such as baseline risk and severity of diseases which influence the effect size,[26] ethical approval and sponsorship) explain differences in the reported effectiveness of c-RCTs, the main project aims to further underscore possible shortcomings, and provide further information and help in the design of studies of complex interventions.

Author affiliations
[1]Institute of General Practice, Goethe University, Frankfurt am Main, Germany
[2]Institute of General Practice and Evidence-based Health Services Research, Medical University of Graz, Graz, Austria
[3]AQUA-Institute for Applied Quality Improvement and Research in Health Care, Göttingen, Germany
[4]Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Graz, Austria

Acknowledgements Paul Glasziou, Sarah Thorning and Elaine Beller extensively discussed the project idea during the research fellowship of Andrea Siebenhofer at the Centre for Research in Evidence-Based Practice, Bond University, Australia in 2013.

Contributors AS was responsible for the conceptual design of the review. The manuscript was drafted by AS and SE, and was revised by CM. SE was responsible for the feasibility check, and AS critically revised the manuscript. Additional statistical analysis for the feasibility check was conducted by AB and GP. The final version of this article has been reviewed and approved by all authors.

Funding The systematic review will be funded by the German Federal Ministry of Education and Research (FKZ: 01KG1504), Germany.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen. Gutachten 2009 des Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen. Koordination und Integration—Gesundheitsversorgung in einer Gesellschaft des längeren Lebens. 2009;16/13770:1–520.
2. Eldridge SM, Ashby D, Feder GS, *et al*. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials* 2004;1:80–90.
3. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328:702–8.
4. Campbell MK, Piaggio G, Elbourne DR, *et al*. Consort 2010 statement: extension to cluster. *BMJ* 2012;345:e5661.
5. Taljaard M, Weijer C, Grimshaw JM. The Ottawa Statement on the ethical design and conduct of cluster randomised trials: précis for researchers and research ethics committees. *BMJ* 2013;346:f2838.
6. Craig P, Dieppe P, Macintyre S, *et al*. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;337:979–83.
7. Djulbegovic B, Kumar A, Glasziou PP, *et al*. New treatments compared to established treatments in randomized trials. *Cochrane Database Syst Rev* 2012;10:MR000024.
8. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). General methods: Version 4.1 (28.11.2013). Inst. für Qual. und Wirtschaftlichkeit im Gesundheitswes. 2013:1–202. https://www.iqwig.de/download/IQWiG_General_Methods_Version_4-1.pdf
9. Taljaard M, McGowan J, Grimshaw JM, *et al*. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. *BMC Med Res Methodol* 2010;10:15.
10. Eldridge S, Ashby D, Bennett C, *et al*. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ* 2008;336:876–80.
11. Institute for Quality and Efficiency in Health Care. Institute for Quality and Efficiency in Health Care. https://www.iqwig.de/
12. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. [Last updated March 2011]. 2011. http://handbook.cochrane.org/
13. Froud R, Eldridge S, Diaz Ordaz K, *et al*. Quality of cluster randomized controlled trials in oral health: a systematic review of reports published between 2005 and 2009. *Community Dent Oral Epidemiol* 2012;40(Suppl 1):3–14.
14. Donner A, Klar N. *Design and analysis of cluster randomization trials in health research*. London: Arnold, 2000.
15. Smeeth L, Siu-Woon Ng E. Intraclass correlation coefficients for cluster randomized trials in primary care. *Control Clin Trials* 2002;23:409–21.
16. Adams G, Gulliford MC, Ukoumunne OC, *et al*. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004;57:785–94.
17. Steventon A, Bardsley M, Billings J. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *BMJ* 2012;344:e3874.
18. Griffiths C, Foster G, Barnes N, *et al*. Specialist nurse intervention to reduce unscheduled asthma care in a deprived multiethnic area: the East London Randomised Controlled Trial for High Risk Asthma (ELECTRA). *BMJ* 2004;328:144.
19. Murphy AW, Cupples ME, Smith SM, *et al*. Effect of tailored practice and patient care plans on. *BMJ* 2009;339:b4220.

20. Möhler R, Köpke S, Meyer G. Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare: revised guideline (CReDECI 2). *Trials* 2015;16:1–9.
21. Möhler R, Bartoszek G, Köpke S MG. Proposed Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare (CReDECI): guideline development. *Int J Nurs Stud* 2012;49:40–6.
22. Diaz-Ordaz K, Froud R, Sheehan B, *et al*. A systematic review of cluster randomised trials in residential facilities for older people suggests how to improve quality. *BMC Med Res Methodol* 2013;13:127.
23. Ivers NM, Taljaard M, Dixon S, *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;342:d5886.
24. Turner L, Shamseer L, Altman DG, *et al*. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 2013:11:MR000030.
25. Díaz-Ordaz K, Kenward MG, Cohen A, *et al*. Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. *Clin Trials* 2014;11:590–600.
26. Muth C, van den Akker M, Blom JW, *et al*. The Ariadne principles: how to handle multimorbidity in primary care consultations. *BMC Med* 2014;12:223.

SUPPLEMENTARY FILE 1

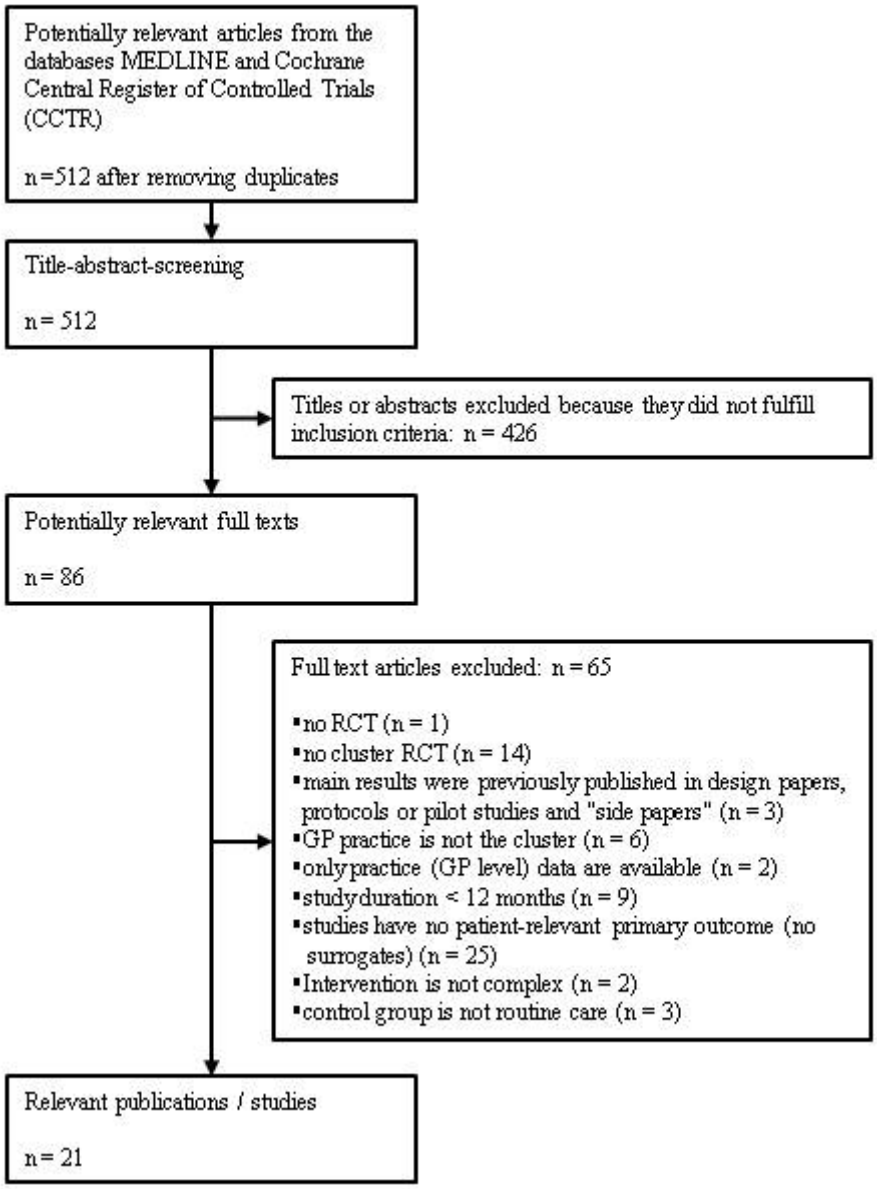

Figure 1: Result of the literature research

Table 1: Excluded studies in full texts with author and reason for exclusion

Publication	**Reason for exclusion**
Arthur 2002	No c-RCT
Baker 2003	No patient-relevant primary endpoints
Bennewith 2002	No patient-relevant primary endpoints
Bergholdt 2012	Study duration < 12 months
Bergholdt 2013	Study duration < 12 months
Bischoff 2012	No c-RCT
Boyd 2010	No patient-relevant primary endpoints
Burns 1998	No c-RCT
Campbell 1997	No c-RCT
Campbell 1998	No c-RCT

Publication	Reason for exclusion
Community Pharmacy Medicines Management Project Evaluation 2007	No c-RCT
Cranney 1999	Study duration < 12 months
Davies 2008	No patient-relevant primary endpoints
de Groot 2007	No patient-relevant primary endpoints
Dean 2014	No c-RCT
Donnan 1990	No c-RCT
Eccles 2000	Design papers, protocols or pilot studies
Eccles 2002	Control group is not routine care
Elwyn 2004	Data is only available for practices (on a general practitioner level)
Fitzmaurice 1996	Design papers, protocols or pilot studies
Flottorp 2003	Data is only available for practices (on a general practitioner level)
Gallo 2013	Design papers, protocols or pilot studies
Glasgow 2003	No patient-relevant primary endpoints
Goldfeld 2011	Setting and cluster are general practices
Harting 2006	No patient-relevant primary endpoints
Hicks 2008	No patient-relevant primary endpoints
Imperial Cancer Research Fund General Practice Research Group 1994	No c-RCT
Kaner 2013	Control group is not routine care
Kendrick 1999	No patient-relevant primary endpoints
King 2002	Study duration < 12 months
Kinmonth 1998	No patient-relevant primary endpoints
Langham 2002	No patient-relevant primary endpoints
Lester 2003	No complex intervention
Lester 2007	Study duration < 12 months
Liaw 1996	Study duration < 12 months
Maclean 2009	Setting and cluster are general practices
Mitchell 2005	No patient-relevant primary endpoints
Montgomery 2000	No patient-relevant primary endpoints
Moore 2003	No patient-relevant primary endpoints
Morgan 2013	Study duration < 12 months
Morrell 2009	No patient-relevant primary endpoints
Murchie 2003	No patient-relevant primary endpoints
Murchie 2004	No c-RCT
Oakeshott 2000	No complex intervention
Pearl 2003	No patient-relevant primary endpoints
Pierce 2000	No c-RCT
Pill 1998	No patient-relevant primary endpoints
Powell 2004	Setting and cluster are general practices
Premaratne 1999	No c-RCT
Putnam 1989	Setting and cluster are general practices
Qureshi 2012	Setting and cluster are general practices
Reiff-Hekking 2005	Setting and cluster are general practices
Ridsdale 1997	No patient-relevant primary endpoints
Russell 1993	No patient-relevant primary endpoints
Schroeder 2005	Study duration < 12 months
Shum 2000	Study duration < 12 months
Smeeth 2003	Control group is not routine care
Smith 2004	No patient-relevant primary endpoints
Steptoe 1999	No patient-relevant primary endpoints
Thapar 2002	No patient-relevant primary endpoints
van Limpt 2011	No patient-relevant primary endpoints
Vetter 1984	No c-RCT
Winters 1997	No c-RCT
Woodfine 2011	No c-RCT
Wright 1998	No patient-relevant primary endpoints

SUPPLEMENTARY FILE 2

Additional information on the ICCs

After requesting additional information on the actually obtained ICCs in order to compare them to those assumed in the sample size calculation, we had both values for 7, or one third of the studies. Two of those were studies that showed no significant treatment effect, while two showed a "partial effect" and three showed a significant treatment effect on all of their respective primary outcomes (which explains the 100% availability of ICC pairs for this particular subgroup). Altogether, 9 assumed and actual ICC pairs for various patient-relevant primary endpoints were available for these 7 publications. In half of the pairs, the assumed ICCs were smaller than actually obtained ICCs (see Table 1)

Table 1: Number of expected and achieved intra-cluster correlation coefficients

Author	Primary patient-relevant outcome	ICC - expected	ICC - achieved
Kennedy 2013[1]	Generic health-related quality of life	0.05	0.031
Steventon 2012	Proportion of people admitted to hospital within the 12 month trial period	0.001	0.017
Gensichen 2009[1]	Depression symptoms	0.1	0.061
Griffiths 2013	Percentage of participants attending for unscheduled asthma care	0.05	-0.0056
Murphy 2009	Changes in physical health status	0.001	0.076
Murphy 2009	Changes in mental health status	0.001	0.054
Murphy 2009	Admissions to hospital	0.006	0.017
Metzelthin 2013[1]	Disability	0.05	0.4
Elley 2003[1]	Vitality	0.05	0.01

[1] data obtained from authors

SUPPLEMENTARY FILE 3

RESULTS OF THE FEASIBILITY PROJECT

Selection of trials and data collection for the feasibility project

The abstracts and titles were independently screened by two reviewers (SE and AS) and the full texts by SE only. However, when the suitability of a publication was in doubt, it was double-checked by AS, and disagreements between the reviewers were resolved by discussion. Consensus on inclusion or exclusion was reached for both abstracts and full text reviews.

Results of the feasibility project

Identification of studies

After removing duplicates, 512 papers were identified during the initial search. Of those, 426 were excluded in the abstract screening process. The full texts of the remaining 86 studies were screened, and a further 65 studies excluded as a result; 21 papers ultimately fulfilled the inclusion criteria (see Figure 1 at the end of the text). Full texts of the excluded studies and information on the author and reasons for exclusion are shown in Supplementary File 1.

Eighteen authors were contacted for further information on the ICCs used in their studies; three authors had already provided the necessary information in their publications. Of those 18, four then provided the relevant information (see Supplementary File 2).

Description of the studies

The reported criteria in the studies were grouped thematically (see Table 1 at the end of the text).

The 21 included studies were published between 1995 and 2013; six of them were published in 2013 alone. Eight of the studies were from the United Kingdom, followed by the Netherlands and the USA (three studies each), and most of them were published in the British Medical Journal (13 studies).

The most common groups of patients were patients with respiratory and mental diseases (six studies each). Patients with diabetes mellitus were examined in four studies and five involved elderly persons.

All studies examined complex interventions but of different levels of complexity. Most of them dealt with interventions that aimed to improve outcomes by means of a multifaceted program. They also differed in terms of the persons delivering the intervention, who were either general practitioners or specialized nurses (see Table 2 at the end of the text).

Outcome measures

Our analysis revealed that the majority of the studies (67%) could not show an intervention effect on the primary patient-relevant endpoint (see Table 3 at the end of the text). Of the 21 examined studies, 14 could not demonstrate such an effect, while three studies did reveal an intervention effect on the primary patient-relevant endpoint. The feasibility project also identified four studies that had more than one primary outcome and showed effectiveness as well as ineffectiveness, depending on the endpoint (referred to as "partly effective"). As we discovered potential differences in quality between c-RCTs that may to some extent determine whether results come out in favour of a complex intervention, we decided to exclude these studies from the review. Exactly which interventions showed an effect and the size of these on primary patient-relevant outcomes are described in Table 5 at the end of the text.

Differences in study quality between studies with and without an intervention effect

As far as general information is concerned, the criteria "patient consent" and "ethical approval" were reported in the majority of trials (86 and 71%, respectively) that were unable to show an effect on the primary outcome. Less frequently, the details on consent of clusters (36%), publication of a study protocol (43%), and trial registration (36%) were provided. In comparison, studies that found a significant intervention effect more frequently provided four of the five listed criteria (see Table 4 at the end of the text).

Some quality criteria concerning the sample size calculation were provided in 86% of studies that showed no superiority. However, consideration of the ICC and involvement of the cluster size in the sample size calculation were described less frequently (64% and 50%, respectively) and only few studies (21%) provided information on whether the cluster size was identical at baseline. In studies

showing a significant effect on the primary endpoint, this information was provided in full, with the exception of the identical cluster size at baseline (see Table 4 at the end of the text).

The method of randomization was only presented clearly in 64% of the studies that showed no intervention effect but in all studies that demonstrated superiority. However, irrespective of the significance of the primary outcomes, all other criteria in this category (recruitment and identification bias, allocation concealment, blinding (patients and outcomes)) were either reported poorly or not at all (see Table 4 at the end of the text).

In terms of analysis method, most of the studies that showed no intervention effect dealt with patient drop-outs (86%) and clusters (71%), performed ITT analyses (71%) and generally accounted for clustering in the analysis (86%). In studies showing an intervention effect on primary outcomes, 67% presented information on cluster drop-outs, and all other quality criteria that were mentioned were reported completely (see Table 4 at the end of the text).

Limitations

No conclusions can be drawn as to whether or not c-RCTs conducted in a general practice setting more often fail to show the effectiveness of a complex intervention due to methodological shortcomings. Our feasibility test did not enable us to rule out that intervention effects were simply lacking, i.e., an intervention was just not effective or not effective enough. But despite our limited sample, we were able to point out some aspects which will be investigated systematically in the planned full review. Secondly, we must consider that the included studies may reflect selection bias, as we only searched for c-RCTs in certain types of journal - the aim of the full review is to correct for this and to achieve an unbiased view. Thirdly, the limited number of included c-RCTs did not allow us to prioritize from among different CONSORT items and to ascertain the methodological quality of the trial: e.g. methods after trial commencement (the way in which an intervention is delivered and implemented and whether or not the investigators defined its fidelity) may be more important than whether the term "cluster randomized trial" appeared in the title. Fourthly, our

feasibility trial did not comprehensively examine methodological shortcomings that concern the gradual development and evaluation of a complex intervention. Thus it did not attempt to answer such questions as (1) whether a study had a sound theoretical foundation, (2) whether the piloting of the intervention components, outcomes and processes justified confidence in the feasibility of the project, (3) whether the effectiveness of the intervention had been appropriately evaluated, and (4) whether process evaluation had been well planned a priori. The full review will therefore have to take these more specific aspects into consideration by examining the framework of the c-RCT, the fidelity of the intervention, and barriers and facilitators to its implementation.

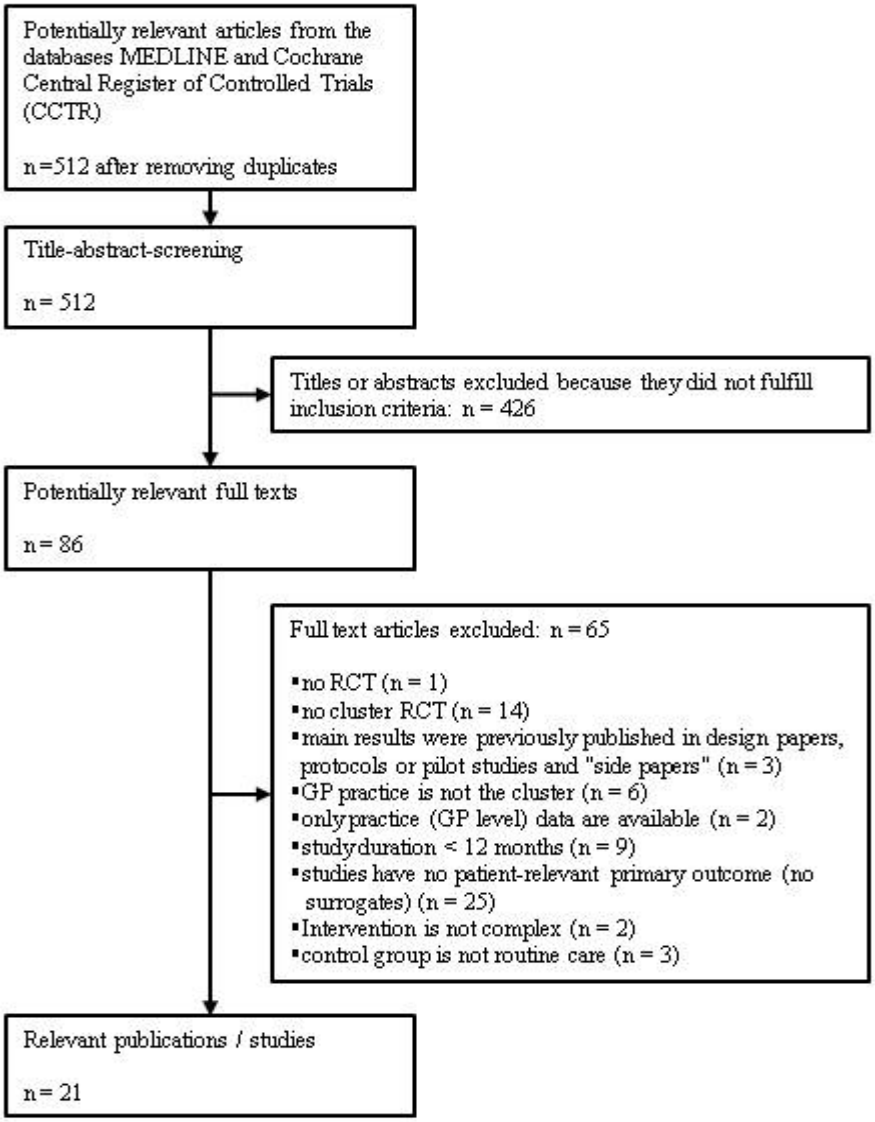

Figure 1: Result of the literature research

Table 1: Data extraction form

General information	Title, authors, journal, date of publication, country of publication, funding/conflict of interests (according to the author) and c-RCT evident from title
Study characteristics	Study design, objective (including target population/ health condition of the subject group), setting of the study, number of participating practices, cluster and cluster size (number of clusters screened, randomized and analysed), patients (number of patients screened, included, analyzed and lost to follow up), patient-relevant primary endpoint (s), not patient-relevant primary endpoint (s), patient-relevant secondary endpoint (s), not patient-relevant secondary endpoint (s)
Baseline Data	Baseline characteristics (cluster and patients), age, ethnicity, sex of the patients, disease-orientated information, inclusion criteria (cluster and patients)
Intervention Data	Run-in phase, contents of the intervention and control groups, recruiting period, follow-up period, observation period for the intervention and control groups
Outcome Data	Intervention effects on primary endpoint(s) including significance level, intervention effects on secondary endpoint(s), intra-cluster correlation coefficient (Are ICCs calculated for the primary endpoint or is information available on the effect of the design?), results of sub-group analyses, p-values (for baseline data)
Quality of the studies - general	Ethical approval, trial registration, sample size calculation method, recruitment method (cluster and patient level), consent ((clusters and patients), before or after the randomization of the practice), publication of the study protocol, involvement of the cluster in the 1st sample size calculation and 2nd. analysis, generalizability of the results for cluster and patients (according to the author), identical cluster size at baseline, recruiting/identification bias (possibility of bias adopted according to Eldridge 2008: not possible, unclear or unlikely)
Quality of the studies - risk of bias	Appropriate randomization method (acceptable: random number table, computer-generated random numbers, minimization, inappropriate: coin flip), acceptable allocation concealment (central allocation and sequentially numbered, opaque, sealed envelopes) blinding (open, blind, double-blind), dealing with drop-out (clusters and patients), intention to treat analysis (ITT), other potential bias (according to the author)
Authors' own interpretation/ explanation	Extraction of reasons why their studies did not show a positive effect e.g. loss to follow up, issues related to recruitment, adherence and data collection (outcomes).

Table 2: Description of the included studies

First author and year	Journal	Publication country	Target population / health condition	Aim / Objective
Bould 2013	Journal of General Internal Medicine	USA	Elderly people	To assess patients' functional health when guided care teams provide proactive, coordinated, comprehensive care
Byng 2004	British Journal of General Practice	United Kingdom	Mental illness	To determine patient satisfaction with care and patient perceptions with regard to unmet needs in the Mental Health Link program designed to improve communication between the teams and systems of care within general practice
Cartwright 2013	British Medical Journal	United Kingdom	COPD, diabetes, heart failure	To assess the effect of second generation, home-based tele-health on health-related quality of life, anxiety, and depressive symptoms
Elley 2003	British Medical Journal	New Zealand	Elderly people	To assess the long-term effectiveness of the green prescription programme on quality of life. The program provides advice on physical activity in a general practice setting
Gallo 2007	Annals of Internal Medicine	USA	Elderly people	To test whether an intervention to improve depression care can influence the risk of death
Gensichen 2009	Annals of Internal Medicine	Germany	Depression	To determine the effects of case management provided by health care assistants in small primary care practices on depression symptoms
Griffiths 2004	British Medical Journal	United Kingdom	Asthma	To determine the influence of specialist asthma nurses in a deprived multi-ethnic area on the percentage of participants attending a practice for unscheduled asthma care, and the time to first attendance for unscheduled asthma care the year after the intervention
Guldin 2013	Family Practice	Denmark	Relatives of patients after death by cancer	To test whether the implementation of a bereavement management program improves the general practitioner's ability to identify complicated grief and provide clinical care
Jarmann 2002	British Medical Journal	United Kingdom	Parkinson`s disease	To determine the effects of community-based specialist nurses on specific measures of health and patient well-being
Jellema 2005	British Medical Journal	Netherlands	Unspecific low back pain	To compare the differences between a minimal intervention strategy and usual care on the treatment of (sub) acute lower back pain on functional disability
Kennedy 2013	British Medical Journal	United Kingdom	Diabetes, COPD, irritable colon	To determine the effectiveness of an intervention to enhance self management support for patients with chronic conditions on generic health-related quality of life
Kerse 1999	British Medical Journal	Australia	Elderly people	To establish the effect of an educational intervention for general practitioners on the functional status of patients
Kinnersley 1999	Family Practice	United Kingdom	Dermatologic, orthopaedic, gynaecologic, rheumatic, ophthalmologic diseases	To describe whether in-house referral is practicable and acceptable for patients and whether it improves patient health outcomes and management in primary care
Metzelthin 2013	British Medical Journal	Netherlands	Elderly people	To evaluate the effect of an interdisciplinary primary care approach on disability
Murphy 2009	British Medical Journal	Ireland	Coronary heart disease	To test the effectiveness of a complex intervention designed within a theoretical framework on the rate of admissions to hospital and physical and mental health status

First author and year	Journal	Publication country	Target population / health condition	Aim / Objective
Olivarius 2001	British Medical Journal	Denmark	Diabetes	To assess the effect of a multifaceted general practice intervention on overall mortality and the patient's disease
Rubenstein 2006	Journal of General Internal Medicine	USA	Depression	To evaluate the effects of EBQI (evidence-based quality improvement) - a method for practices to self-improve depression care performance - on depression care and outcomes
Steventon 2012	British Medical Journal	United Kingdom	Diabetes, COPD, heart failure	To assess the effect of home-based tele-health interventions on the rate of admissions to hospital
Van Marwijk 2008	British Journal of General Practice	Netherlands	Depression	To test the effects of an intervention program that aims to improve the identification, diagnosis, and treatment of depression
Walters 2013	British Medical Journal open	Australia	COPD	To assess the benefits of telephone-delivered health mentoring on health-related quality of life
White 1995	British Medical Journal	United Kingdom	Asthma	To test the effects on classic patient symptoms of feeding back information on patients' asthma to primary care teams

Table 3: Effects on primary patient-relevant outcome (most recent studies first)

Studies	Effect on primary patient-relevant endpoint(s)[1]	Primary patient-relevant endpoint(s)[2]
Boult 2013	↔	Patients' functional health (-)
Cartwright 2013	↔	Treatment effectiveness (-) Treatment efficacy (-)
Guldin 2013	↔	Bereaved relatives' score (-) Relative's number of contacts with general practice (-)
Kennedy 2013	↔	Generic health-related quality of life (-)
Metzelthin 2013	↔	Disability (-)
Walters 2013	↔	Health-related quality of life (-)
Van Marwijk 2008	↔	Montgomery Åsberg Depression Rating-Scale (-) PRIME-MD Scores (-)
Rubenstein 2006	↔	Appropriate depression treatment (-) Recovery from depression (after 12 months) (-)
Jellema 2005	↔	Functional disability (-)
Byng 2004	↔	Patient satisfaction with care (-) Patient perceptions on unmet need (-)
Olivarius 2001	↔	Overall mortality (-) Incidence of diabetic retinopathy (-) Myocardial infarction (-) Stroke in patients without symptoms at baseline (-)
Kerse 1999	↔	Functional status (-)
Kinnersley 1999	↔	Patient satisfaction (-) Health status (-) Management in primary care before and after referral (-)
White 1995	↔	Classic symptoms (-)
Steventon 2012	↑	Proportion of people with an inpatient admission to hospital within the 12 month trial period (+)
Gensichen 2009	↑	Depression symptoms (+)
Griffiths 2004	↑	Percentage of participants attending for unscheduled asthma care (+) Time to first attendance for unscheduled asthma care in the year after the intervention (+)
Murphy 2009	↑/↔	Admissions to hospital (+) Changes in physical and mental health status (-)
Gallo 2007	↑/↔	Mortality: All patients with depression and major depression disorder (+) Clinically significant minor depression and patients without depression (-)
Elley 2003	↑/↔	Quality of life: General health, role physical, vitality, bodily pain (+) Physical functioning, social functioning, role emotional, mental health (-)
Jarmann 2002	↑/↔	Measures of health (-) Patient wellbeing (-) Global health question (+)

1 (↑): Upward arrow: Studies showing an intervention effect; (↔): Horizontal arrow: Studies showing no effect; (↑/↔): Studies presenting more than one primary patient-relevant endpoint with an effect on one or more endpoints but not on all of them within one and the same study

2 (+): Superiority of intervention group for a patient-relevant endpoint demonstrated; (-): No superiority of intervention group for a patient relevant-endpoint demonstrated

Table 4: Differences in study quality between studies with and without an intervention effect on the primary outcome

Study Information	Studies without intervention effect n=14 (% in brackets)	Studies with intervention effect n=3 (% in brackets)
General information		
Consent (patients)	12 (86)	3 (100)
Consent (cluster)	5 (36)	2 (67)
Ethical approval	10 (71)	2 (67)
Publication of study protocol	6 (43)	2 (67)
Trial registration number	5 (36)	2 (67)
Sample size calculation		
Sample size calculation	12 (86)	3 (100)
Assumed ICC	9 (64)	3 (100)
Involvement of the cluster in the sample size calculation	7 (50)	3 (100)
Identical cluster size at baseline	3 (21)	1 (33)
Randomization and blinding process		
Recruiting-/Identification bias	1 (7)	0 (0)
Adequate randomization method	9 (64)	3 (100)
Adequate allocation concealment	2 (14)	1 (33)
Blinding (patients)	4 (29)	1 (33)
Blinding of outcomes assessors	7 (50)	1 (33)
Analysis		
Dealing with drop-out (patients)	12 (86)	3 (100)
Dealing with drop-out (cluster)	10 (71)	2 (67)
ITT	10 (71)	3 (100)
Involvement of cluster in the analysis	12 (86)	3 (100)

Table 5: Which interventions showed an effect and the size of the effects on primary patient-relevant outcomes

Studies	Intervention effects on primary patient-relevant outcomes (with significance level)
Boult 2013	Patients' functional health: Physical Health: Difference Guided Care/Usual Care: -1.31 (CI: -3.02-0.41) Mental Health: Difference Guided Care/Usual Care: 1.05 (CI: −1.08-3.12) (adjusted for baseline age, race, sex, education level, financial status, habitation status, HCC score, SF-36 physical and mental health subscales, and satisfaction with health care)
Byng 2004	Patients' satisfaction with care: Adjusted difference between control and intervention at follow-up: -0.01 (CI: -0.21-0.18; P=0.88) (controlling for baseline scores and allowing for clustering of patients within practices) Patients' perception of unmet need: Adjusted difference between control and intervention at follow-up: -0.02 (CI: -0.56-0.51; P=0.94) (controlling for baseline scores and allowing for clustering of patients within practices)
Cartwright 2013	Treatment effectiveness with intention to treat analysis (ITT): No significant differences between the groups for the patient-relevant outcomes quality of life, depression symptoms and anxiety Complete case: $0.480 \leq P \leq 0.904$ Available case (baseline data and data of one other assessment): $0.181 \leq P \leq 0.905$ Treatment efficacy with per-protocol analysis: No significant differences between the groups for the patient-relevant outcomes quality of life, depression symptoms and anxiety Complete case: $0.273 \leq P \leq 0.761$ Available case (baseline data and data of one other assessment): $0.145 \leq P \leq 0.696$
Elley 2003	Quality of life: Difference between groups (adjusted for clustering by medical practice): general health: 4.51 (CI: 2.07-6.95; P=0.000) physical fitness: 7.24 (CI: 0.16-14.31; P=0.045) vitality: 2.30 (CI: 0.03-4.57; P=0.047) bodily pain: 4.01 (CI: 0.78-7.24; P=0.02) physical functioning: 1.23 (CI: -1.35-3.81; P=0.3) social functioning: 0.36 (CI:-3.53-4.26; P=0.9) emotional status: -0.38 (CI: -5.70-4.94; P=0.9) mental health: 0.98 (CI: -0.99-2.95; P=0.3)
Gallo 2007	Mortality: Hazard ratio for intervention effects (includes terms for baseline age, sex, education, smoking, cardiovascular disease, stroke, diabetes, cancer, cognition, and suicidal ideation): All patients with depression: 0.67 (CI:0.44-1.00) Major depression disorder: 0.55 (CI: 0.36-0.84) Clinically significant minor depression: 0.97(CI: 0.49-1.92) Patients without depression: 1.14 (CI: 0.84-1.53)
Gensichen 2009	Depression symptoms: Mean difference (P-value based on a 2-level linear mixed model for respective outcomes (T1 and T2), adjusted for intracluster correlation and baseline depression): -1.41 (CI: -2.49 to -0.33; P=0.014)
Griffiths 2004	Percentage of participants attending for unscheduled asthma care: Adjusted odds ratio (with clustering): 0.61 (CI: 0.38-0.99) Adjusted odds ratio (without clustering): 0.62 (CI: 0.38-1.01) Time to first attendance for unscheduled asthma care in the year after intervention: Hazard ratio: 0.73 (CI: 0.54-1.00)
Guldin 2013	Bereaved relatives' score - depression: Mean score, intervention group: 7.85 (CI: 6.53-9.17) Mean score, control group: 8.84 (CI: 7.41–10.28) Bereaved relatives' score - grief symptoms: Mean score, intervention group: 14.73 (CI:13.14-16.32)/ Mean score, control group: 15.57 (CI: 13.77-17.38) Relatives' number of contacts with general practice: Contact frequencies with GPs: Corresponding rate ratio: 0.92 (CI: 0.72–1.17); P=0.50 Out-of-hours contacts with GPs: Corresponding rate ratio: 0.55 (CI: 0.29-1.06); P=0.07

Studies	Intervention effects on primary patient-relevant outcomes (with significance level)
Jarmann 2002	Measures of health: Bone fracture during study: Odds Ratio: 1.20 (CI: 0.85-1.69); P=0.31 Mortality (2 years): Hazard ratio: 0.91 (CI: 0.73-1.13) P=0.38 Mortality (4 years): Hazard ratio: 0.89 (CI: 0.76-1.03); P=0.12 Patient wellbeing: Euroqol: Difference: -0.02 (CI: -0.06-0.02); P=0.30 PDQ-39 summary index: Difference: 0.47 (CI: -2.72-3.66); P=0.77 Global health question: Difference: -0.23 (CI: -0.40 to -0.06); P=0.008
Jellema 2005	Functional disability: Mean difference (adjusted for baseline values): 0.25 (CI: −0.77-1.28)
Kennedy 2013	Generic health-related quality of life: Adjusted mean difference (adjusted for model factors and covariates): -0.00 (CI: -0.02-0.01) Effect size (Adjusted mean difference (intervention minus control) divided by standard deviation in practice): -0.01 (CI: -0.05-0.04); P=0.72 P value for interaction with condition group (P value for test of whether intervention effect varies by disease condition): 0.31
Kerse 1999	Functional status: Mean effect size: 2.10 (CI: -0.94-5.1); P= 0,175 (All analyses were controlled for general practitioner billing status and effect of cluster design)
Kinnersley 1999	Patient satisfaction (mean): Intervention group (referred immediately to secondary care): 80.7 (SD: 11.1) Intervention group (not referred): 78.5 (SD: 12.2) Control group: 79.2 (SD: 10.3) Health status (mean): Intervention group (referred immediately to secondary care): 64.4 (SD: 33.5) Intervention group (not referred): 77.1 (SD: 27.9) Control group: 67.9 (SD: 29.6) Management in primary care before and after referral (mean): Intervention group (referred immediately on to secondary care): 0.25 (SD: 0.5) Intervention group (not referred): 0.56 (SD: 0.69) Control group: 0..36 (SD:0.65)
Metzelthin 2013	Disability (after 12 months): Mean difference (adjusted for age, sex, education, and significant differences at baseline (frailty and disability)): 0.47 (CI: -0.81 to 1.76); P=0.47
Murphy 2009	Admissions to hospital: Mean difference: -0.15 (CI: -0.01 to -0.29); P= 0.03 (ICC: 0.017) Changes in physical health status: Mean difference: -0.78 (CI: -2.58-1.03); P=0.39 (ICC: 0.076) Changes in mental health status: Mean difference: 0.02 (CI: -2.40 -2.35); P= 0.98 (ICC: 0.054)
Olivarius 2001	Overall mortality: P=0.82 Incidences of diabetic retinopathy: Odds ratio: 0.90 (KI: 0.53-1.52); P=0.69 Myocardial infarction: Odds ratio: 0.65 (KI: 0.31-1.35); P=0.25 Stroke in patients without these outcomes at baseline: Odds ratio: 0.89 (KI: 0.39-2.01); P=0.77
Rubenstein 2006	Appropriate depression treatment and recovery from depression (after 12 months): Effect size: 0.03; P=0.77 Intervention group (Mean): 45.6 (CI: 37.8-53.5) Control group (Mean): 47.0 (CI: 42.7-51.3) (All regressions controlled for covariates (age, sex, completion of high school, household wealth, timing of enrolment, ethnicity, count of chronic diseases, marriage, alcohol use, dysthymia) and baseline values of the dependent variable)
Steventon 2012	Proportion of people with an inpatient admission to hospital within the 12-month trial period: Unadjusted odds ratio: : 0.82 (CI: 0.70-0.97); P=0.017 Adjusted odds ratio: 0.82 (CI: 0.69-0.98); P=0.026 Combined model odds ratio: 0.82 (CI: 0.69-0.96); P=0.016
van Marwijk 2008	Montgomery Åsberg Depression Rating-Scale: Intervention group (mean): 10.80 (SE 2.85) Control group (mean): 10.09 (SE 2.50) PRIME-MD Scores: Intervention group (mean): 3.23 (SE 1.04) Control group (mean): 3.74 (SE 1.21)

Studies	Intervention effects on primary patient-relevant outcomes (with significance level)
Walters 2013	Health-related quality of life: SGRQ (mean): Intervention group: 41.9 (SD: 18.9) Control group: 40.5 (SD:17.4) SF-36 - Mental health component summary (mean): Intervention group: 50.2 (SD: 11.4) Control group: 50.5 (SD: 10.5) SF-36 - Physical component summary (mean): Intervention group: 38.5 (SD: 10.3) Control group: 38.5 (SD: 9.4)
White 1995	Classic symptoms: Breathlessness at least once a week (mean): Intervention group: 36.0 (SD:14.3) Control group: 35.0 (SD: 10.9);P=0.79 Wheeze at least once a week (mean): Intervention group: 38.0 (SD: 11.7) Control group: 31.0 (SD: 14.4); P=0.19 Cough at least once a week (mean): Intervention group: 49.0(SD:13.9) Control group: 45.0(SD: 12.1); P=0.47 Night waking at least once a week (mean): Intervention group: 27.0 (SD: 9.9) Control group: 23.0 (SD: 11.2); P= 0.39 Any time off work or studies due to asthma (mean): Intervention group: 16.8 (SD: 7.3) Control group: 19.1 (SD: 6.7); P=0.45 At least one severe attack (mean): Intervention group: 49.3(SD: 13.1) Control group: 43.3(SD: 13.2); P=0.3 Breathless on level ground (mean): Intervention group: 41.3 (SD: 17.0) Control group: 48.1 (SD: 13.0);P=0.7 Any attendance at surgery (doctor or nurse) (P=0.96), Regular use of inhaled steroids (P=0.62) and Regular use of inhaled bronchodilator (P=0.78)