BMJ Open

Pilot study to test the feasibility of a trial design and complex intervention on *PRI*oritising *MU*ltimedication in *M*ultimorbidity in general practices (PRIMUM*pilot*)

Christiane Muth,[1] Sebastian Harder,[2] Lorenz Uhlmann,[3] Justine Rochon,[3] Birgit Fullerton,[1] Corina Güthlin,[1] Antje Erler,[1] Martin Beyer,[1] Marjan van den Akker,[1,4,5] Rafael Perera,[6] André Knottnerus,[4] Jose M Valderas,[7] Ferdinand M Gerlach,[1] Walter E Haefeli[8]

To cite: Muth C, Harder S, Uhlmann L, *et al*. Pilot study to test the feasibility of a trial design and complex intervention on *PRI*oritising *MU*ltimedication in *M*ultimorbidity in general practices (PRIMUM*pilot*). *BMJ Open* 2016;**6**:e011613. doi:10.1136/bmjopen-2016-011613

▸ Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2016-011613).

Received 21 February 2016
Revised 24 May 2016
Accepted 22 June 2016

For numbered affiliations see end of article.

Correspondence to
Dr Christiane Muth;
muth@allgemeinmedizin.uni-frankfurt.de

ABSTRACT

Objective: To improve medication appropriateness and adherence in elderly patients with multimorbidity, we developed a complex intervention involving general practitioners (GPs) and their healthcare assistants (HCA). In accordance with the Medical Research Council guidance on developing and evaluating complex interventions, we prepared for the main study by testing the feasibility of the intervention and study design in a cluster randomised pilot study.

Setting: 20 general practices in Hesse, Germany.

Participants: 100 cognitively intact patients ≥65 years with ≥3 chronic conditions, ≥5 chronic prescriptions and capable of participating in telephone interviews; 94 patients completed the study.

Intervention: The HCA conducted a checklist-based interview with patients on medication-related problems and reconciled their medications. Assisted by a computerised decision-support system (CDSS), the GPs discussed medication intake with patients and adjusted their medication regimens. The control group continued with usual care.

Outcome measures: Feasibility of the intervention and required time were assessed for GPs, HCAs and patients using mixed methods (questionnaires, interviews and case vignettes after completion of the study). The feasibility of the study was assessed concerning success of achieving recruitment targets, balancing cluster sizes and minimising drop-out rates. Exploratory outcomes included the medication appropriateness index (MAI), quality of life, functional status and adherence-related measures. MAI was evaluated blinded to group assignment, and intra-rater/inter-rater reliability was assessed for a subsample of prescriptions.

Results: 10 practices were randomised and analysed per group. GPs/HCAs were satisfied with the interventions despite the time required (35/45 min/patient). In case vignettes, GPs/HCAs needed help using the CDSS. The study made no patients feel uneasy. Intra-rater/inter-rater reliability for MAI was excellent. Inclusion criteria were challenging and potentially inadequate, and should therefore be adjusted. Outcome measures on pain, functionality and self-reported adherence were unfeasible due to frequent missing values, an incorrect manual or potentially invalid results.

Conclusions: Intervention and trial design were feasible. The pilot study revealed important limitations that influenced the design and conduct of the main study, thus highlighting the value of piloting complex interventions.

Trial registration number: ISRCTN99691973; Results.

Strengths and limitations of this study

- This is the first randomised piloting of a complex intervention addressing polypharmacy in primary care.
- The studied complex intervention aims to support an interaction assessment, discover patient preferences and use a computerised decision-support system (AiDKlinik) to prioritise polypharmacy.
- The complex intervention addressed the entire medication use process and included a health care assistant of the general practice to empower patients and to reduce physician's workload.
- The pilot study design allowed critical procedures to be implemented and all materials and instruments for the planned main study on the effectiveness of the complex intervention to be tested.
- The pilot study design demanded considerable effort and doubled the sample size for the feasibility testing of the complex intervention.

INTRODUCTION

Currently, as many as 80% of consultations in primary care involve patients with multiple chronic conditions.[1] Multiple disorders in patients are likely to result in the prescription of a number of different drugs and often in polypharmacy (>4 drugs). Polypharmacy is associated with drug underuse, particularly in older people,[2 3] and also poses a substantial risk for adverse drug reactions (ADR) and non-adherence, possibly leading to hospitalisation, cognitive impairment, falls, increased mortality and an increase in healthcare costs.[4–7] About 60% of drug-related hospitalisations are due to inappropriate prescriptions, and about 20% to non-adherence.[8 9] At least half of these are preventable.[10 11]

Although interventions of proven effectiveness on clinical outcomes are still lacking,[12 13] promising strategies aimed at combating inappropriate polypharmacy exist. A first essential step is to get a comprehensive overview of the patient's current medication and intake habits. This can be accomplished by means of a so-called 'brown bag review', in which patients are invited to bring all their medicines to the practice in their original packaging.[14] Concurrently, patient adherence and hitherto unknown prescriptions from other healthcare providers can be assessed.[15 16] This information is necessary if prescribing is to be improved.[17–19] Second, the use of computerised decision support systems (CDSS) can help ensure appropriate prescribing.[20–23] Third, preconsultation interviews provide an opportunity for healthcare assistants (HCAs) to encourage older patients to tell their physicians about any medication-related problems, thus improving adherence.[24]

On the basis of these strategies, we designed a complex intervention to improve prescribing and adherence in older patients with multimorbidity and polypharmacy in general practice in Germany. We also included HCAs from the participating practices. HCAs receive less training in patient care than nurses and are comparable to certified medical assistants in the USA with regard to education, responsibilities and remuneration.[25 26] HCAs have been repeatedly and successfully included in chronic care interventions in Germany: under the supervision of general practitioners (GPs), they have followed evidence-based protocols and algorithms with fixed interview questions, and have provided self-management support or telephone monitoring for conditions such as osteoarthritis, major depression and chronic heart failure.[27–29]

We tested the feasibility of the intervention in a cluster-randomised controlled pilot study.[30] To improve the study design of the main study, we focused on aspects relating to the recruitment of practices and patients, randomisation procedures, prevention of drop-outs and outcome measures that are relevant for subsequent sample size calculations for the main study.[31 32]

METHODS

Design and participants

We performed a cluster-randomised, controlled pilot study with the general practice as the unit of randomisation. We compared a complex intervention (intervention group) with usual care (control group) with an allocation ratio of 1:1. Treatment allocation was concealed to practices and patients until data collection at baseline had been completed (figure 1[33–58]).

We invited academic teaching practices and GPs who attended the Frankfurt General Practice Day to participate in the study. Inclusion criteria for practices were the provision of primary care within the German statutory health insurance system and that the HCA could access the internet. A random sample of patients (for patient recruitment, see figure 1, icons c-e) fulfilling the following criteria was included: age ≥65 years, ≥3 chronic conditions, ≥5 chronic prescriptions, ≥1 practice visit during the past quarter and the ability to fill in questionnaires and participate in telephone interviews. We excluded patients with cognitive impairment (Mini-Mental Status Examination, MMSE <26),[36] because we designed our intervention for cognitively intact patients and did not address caregivers. Further exclusion criteria were a life expectancy ≤6 months, alcohol and drug abuse (based on the GP's assessment).

Intervention and control treatment

Intervention group

The PaTplot[59] (figure 1) shows the elements of the complex intervention. It consists of a brown bag review and a checklist-based preconsultation interview with the patient that is conducted by the HCA (see online supplementary web–appendix 1), a computer-assisted medication review carried out by the GP and a GP-patient consultation. GPs in the intervention group received practice guidelines for older patients,[35] and the complex intervention was implemented at their practice on a single occasion.

Control group

GPs in the control group also received the practice guidelines for older patients,[35] but continued with usual care.

Measurements

Feasibility of the study

The pilot study aimed to test all procedures, materials and instruments for their suitability for use in the main study.[31 32] The achievement of recruitment targets, the balance of cluster sizes, treatment allocation and baseline characteristics in both groups, and reasons for non-participation and loss to follow-up of patients were examined.

Feasibility of outcome measures

Medication appropriateness index (MAI): As a potential primary outcome to be used in the main study, we tested the MAI, because it is widely accepted that it focuses on patients rather than drugs and diseases.[60] This fitted in well with our holistic intervention which was aimed more at optimising medication prescriptions than on reducing the number of prescriptions per se. The MAI consists of 10 items: (1) indication for the drug, (2) efficacy for the condition, (3) correctness of dosage, (4)

correctness of directions, (5) practicality of the directions, (6) drug–drug interactions, (7) drug-disease interactions, (8) unnecessary drug duplications, (9) correctness of treatment duration and (10) cost.[61 62] Item (10) was omitted because variable discount contracts between pharmaceutical and statutory health insurers preclude cost comparisons in Germany. Items (1) to (9) were rated for each prescription on a three-point Likert scale ('1' represented appropriateness, '3' inappropriateness and '2' a middle rating of hardly appropriate). Operational definitions and explicit instructions were determined a priori for each index item. An experienced clinical pharmacologist (SH) coded the MAI following a blinded chart review based on the GP's prescriptions, multimorbidity (diagnoses, Cumulative Illness Rating Scale—CIRS)[33 34] (figure 1, icon f) and ADR symptoms (figure 1, icon h). MAI ratings were transformed by subtracting 1 from the original rating, resulting in values ranging from '0' (best rating) to '2' (worst rating), and adding them up to give an MAI score ranging from 0 to 18 per prescription. MAI sum scores across the entire medication regimen of the patient were calculated and the differences in the MAI sum scores between baseline (T0) and T1 (T1-T0) resp. T2 (T2-T0) were determined with lower MAI scores denoting better prescribing appropriateness. A negative value for T1-T0 or T2-T0 therefore reflected an improvement in prescribing quality. Reliability of the MAI: 6 months after T2, the clinical pharmacologist (SH) received a sample of medication reviews for a second rating (blinded to the results of the first) to determine intra-rater reliability. To explore the benefit of a second independent MAI rating, another experienced clinical pharmacologist, blinded to the results of SH, reviewed the same sample to test inter-rater reliability. The sample was randomly drawn from T1 data until the prespecified sample size was achieved.

We also examined health-related quality of life (EQ-5D index),[37] functional status (WHO Disability Assessment Schedule, WHO-DAS II),[38] adherence and related measures, as these are secondary outcomes that may be used in the main trial. We collected data on self-reported adherence according to Morisky (four items resulting in sum scores of 0–4 points, with low scores indicating good adherence) and the Medication Adherence Rating Scale (MARS; five items resulting in sum scores of 5–25 points, with high scores indicating good adherence).[40 42] We also measured the discrepancy between medicines actually taken (reported at patient's interviews) and medicines prescribed (reported by GP). According to Barat *et al*,[63] we calculated (1) the drug score (DS=number of drugs reported by the patients/number of drugs reported by the GP), (2) the dose score ($DoS=d_1(a_1)+d_2(a_2)+d_3(a_3)+\ldots/n$, where d_i is the drug used by the patients (value 0 or 1), n is the number of drugs in the GP's report, and a_i is the dose-deviation rate calculated by dividing the patient's reported daily dose with the daily dose reported by the GP) and (3) the regimen score ($RS=d_1(b_1)+d_2(b_2)+d_3(b_3)+\ldots/n$, where b_i is the regimen-deviation rate calculated by dividing the patient's reported daily intake frequency (once daily, twice daily, etc.) with the corresponding frequency reported by the GP). Scores outside an interval of 0.8–1.2 were considered to be divergent.[63] Adherence-related measures were complexity of medication (total number of prescriptions, number of single doses/day, Medication Regimen Complexity Index, MRCI)[64] and Beliefs about Medicines Questionnaire.[41] In proxy of under-treatment, pain intensity was measured by means of a single visual rating scale (VRS). The numbers of days in hospital, deaths and symptoms of side effects were analysed. We determined the differences between baseline (T0) and T1 (T0-T1) resp. T2 (T0-T2).

Feasibility of the intervention

We used mixed methods consisting of brief questionnaires, semistructured interviews and case vignettes (figure 1, icons l-o and 6–10; online supplementary web–appendix 2). All interviews were audio taped, transcribed and analysed according to qualitative description and content analysis techniques:[65 66] The answers were coded by two independent researchers and dissent was resolved by discussion. Results were analysed according to a previously designed coding scheme and rated as 'feasible', 'not feasible' and 'feasible with limitations'.[67] For the analysis of the case vignettes, need for technical support with the CDSS was categorised (none, minor—help was needed to execute a specific procedure and major—help was needed with necessary operations). GPs' case vignettes were also analysed for the number of CDSS modules used and reduction in the number of drugs and inappropriate prescriptions.

Estimations of sample sizes

According to earlier suggestions that 30 patients per group would allow a good estimate of mean and SD,[31 68] we aimed to recruit at least 50 patients for each of the control and intervention groups, resulting in an overall sample size of N=100. With a target size of 5 patients per cluster, we needed to recruit 10 GP practices per group. This sample size also allowed the estimation of the intracluster correlation coefficient (ICC) that would be required to support the sample size calculation for the main study.

Inter-rater reliability in the MAI assessment ranged from 0.47 to 0.99, and intra-rater reliability from 0.70 to 0.96.[61 62 69–73] Since a less than moderate κ would be inacceptable in our trial, we assumed the null hypothesis value to be 0.4. With an estimated proportion of 0.3 positive ratings, a two-tailed test and 90% power, we therefore needed N=255 prescriptions to detect a κ of 0.6 (95% CI 0.5 to 0.7).[74 75]

Statistical analysis

For all outcomes, the primary analysis took place according to the intention-to-treat principle. The primary

Timeline	Intervention	Control
Practice recruitment	(1) a b	(1) a b
Patient recruitment	c d e	c d e
Baseline (T0)	f g h	f g h
Cluster-randomisation of practices	i	i
Allocated treatment: 5 weeks after baseline	(2) j (3) (4) (5) k l m	k
1st Follow up (T1): 6 weeks after intervention	f g h (6)	f g h
2nd Follow up (T2): 12 weeks after intervention	f g h	f g h
Individual interview at practice site by a study assistant: after completion of T2 data collection	(7) (8) n (9) (10) o	

Material and methods

(1) **Investigator training**: 90 minutes of education and training for the general practitioner (GP) and one health care assistant (HCA) from the practice on the study protocol, and ethical considerations. GPs were trained in the assessment of patients by means of the Cumulative Illness Rating Scale (CIRS),[33;34] and the recommended treatment standard, a guideline on pharmacotherapy for the aged in general practice.[35] HCAs were trained in performing the Mini Mental Status Exam (MMSE)[36] and both, HCAs and GPs, in data documentation.

a **Documentation of GP and practice data**: after providing written consent, the GP documented his/her socio-demographic parameters (date of birth, sex, clinical experience (ys), duration of work at this practice site, clinical (sub-)specialisation) and practice characteristics (single handed vs. group practice, practice location (urban / suburban / rural) and panel size (in Germany, panel size is expressed as the number of patient registrations in a practice over a 3 month period)

b **Documentation of HCA data**: after providing written consent, the HCA documented date of birth, sex, school leaving certificate, vocational and further qualifications, duration of work as HCA, and level of employment

c **Case finding in practices**: GPs generated a screening list with the practice computer (parameter statistics, patients aged ≥65 ys. in descending order by cost), removed patients from the list that did not fulfill the inclusion criteria, and sent the remaining list of at least 30 patient ID's to the study centre pseudonymously

d **Selection of potentially eligible patients**: at the study centre, the first 5 IDs were removed, from the remaining, 10 ID's were selected by Microsoft EXCEL© random numbers and this random list of patient IDs was sent to the practice

e **Patient recruitment**: GPs invited the patients on the list to participate in the study on their appearance in the practice. After patients had given their written informal consent, HCA used the MMSE on them. Patients, who fulfilled inclusion criteria were recruited for the study until 5 patients / practice had been reached. Patient registration forms on age, gender, in- and exclusion criteria, MMSE scores, contact details for telephone interviews was sent to the study centre.

f **Paper-based case report form (CRF)** on each patient completed by the GP and HCA covered sex, year of birth, insurance status, and home care situation, current diagnoses, medication in accordance with the practice documentation, height, weight, laboratory results (sodium, potassium, serum creatinine if available), co- and multimorbidity (CIRS),[33;34] no. of hospital stays, days in hospital.

g **Paper-based questionnaire completed by patients** consisted of: EQ-5D index (quality of life),[37] WHO Disability Assessment Schedule (WHO DAS-II),[38] Medication Adherence Rating Scale (MARS),[40] Beliefs about Medicine Scale (BMQ),[41] pain intensity during the past week (a single Visual Rating Scale, 10 cm, equidistant scaling: no pain - mild pain - moderate pain - severe pain - extreme pain)

h **Standardised telephone interview** of study assistants with patients on following data: socio-demographic, medication actually taken incl. OTC, herbal and nutritional products (trade name, dosage, and National Drug Code, if possible), technical problems with intake, structured symptoms of side effects (dizziness, dyspnea, tachycardia / palpitations, nausea / vomiting, abdominal pain, bleeding diathesis, difficulties urinating, ankle oedema; frequency expressed as occurrence on one day / several days / almost every day during past two weeks), no. of and reasons for falls, days of hospitalisation during the past 6 months, patient reported adherence to medicine,[42] Geriatric Depression Scale[43;44]

i **Cluster-randomisation:** the first patient from each practice served as the basis for randomisation of this practice as either control or intervention practice. Randomisation was carried out at the study centre. All the patients registered thereafter were dealt with according to practice status (control or intervention), which was assigned in an allocation ratio of 1:1 using a block randomisation procedure of variable block length (4 - 6 - 6 - 4). The allocation sequence was generated by an external researcher using the random number generator of Microsoft EXCEL© and kept confidential to the study team until the end of the study.

Intervention

(2) **Intervention training**: GPs and HCAs received 90-120 minutes of interactive intervention training in the use of the CDSS and the checklist either at the study centre, or at their practice making use of case examples.

j **Checklist-based pre-consultation interview** of patients by HCA: The interview was supported by the **Medication Monitoring List** (**MediMoL**) which consisted of 21 questions on medication-related problems such as symptoms of potential adverse drug reactions (ADR),[45-49] potential under-treatment of pain, adherence problems, (technical) problems in drug administration, information needs, and patient's cross-disease 'universal' health preferences; structured answers in accordance with the traffic light pattern (response categories red for "emergency", orange for "potentially serious", yellow for "potentially important", and green for "no problem") and free text answers (online supplementary web–appendix 1).

(3) **Brown bag review** by HCA: appointed patients brought all his/her medication in its original packaging to the practice for medication reconciliation by the HCA.

(4) The **Computerised decision support system** ("AiDKlinik®") consisted of (i) a drug information system covering all medicines actually marketed in Germany, (ii) pre-scription platforms (mediboxes) to compile the medication documented in the practice software (medibox 1: "practice medication"), the patients' medicines at home (medibox 2: "home medication"), and a third combining the two to assess discrepancies (medibox 3: "coordinated medication"), and (iii) decision support modules on drug interactions, duplicate prescriptions, drug dosing in renal failure, incompatibilities of parenteral drugs, and existing discount contracts to automatically assess the content of each medibox.[50-56] The (iii) modules provided weighted alerts inspired by the traffic light pattern and were accompanied by a short explanation of identified risks and suggestions for its clinical management.
The HCA entered the patient's basic data (e.g. date of birth, sex) and all currently prescribed medicines as documented in the practice software into medibox 1. At the same time as the brown bag review, the HCA entered all medicines into medibox 2. The GP then scrutinised the content of medibox 1 and 2 for agreement and ambiguities and critically examined the alerts issued by the software while taking into consideration medication-related problems and patient preferences as identified by MediMoL. With this information, the GP then adjusted and harmonised the treatment schedule, and saved it in medibox 3.

(5) **Physician-patient-consultation** on medication related problems based on j, (3), (4)

k Treatment in accordance with the **recommended standard** for older patients.[35]

Feasibility testing of the intervention

l **Short questionnaire completed by GPs after each intervention:** required time (self estimation) and overall satisfaction with the intervention (1 item, 6-point Likert scale from 1="very satisfied" to 6="very dissatisfied")

m **Short questionnaire completed by HCAs after each intervention:** required time (self estimation) and overall satisfaction with the intervention (1 item, 6-point Likert scale from 1="very satisfied" to 6="very dissatisfied")

(6) **Semi-structured telephone interviews of patients** by a study assistant: subsequent to the data collection at T1, patients were asked for the benefits and any sort of discomfort arising from the intervention.

(7) **Individual semi-structured interviews of GPs**: GPs described their experience with the intervention; questions focused on the feasibility of the intervention, its components and the effects of the entire intervention on the practice team.

(8) **Intervention case vignettes with GPs**:[57] After completion of the semi-structured interview, GPs were presented with a hypothetical case example including: (1) a written case description on age, sex, social situation, actual complaints, previously known diagnoses (chronic heart failure, atrial fibrillation, hypertension, hyper-cholesterinaemia, osteoarthritis, osteoporosis), clinical parameters (body mass index=30, blood pressure, heart rate), and laboratory test results; (2) a completed MediMoL; (3) a prepared AID+ data record (medibox 1: 8 drugs; medibox 2: 11 drugs; serum potassium, sodium, creatinine clearance). The GP optimised medication in medibox 3. The interviewer checked the correct use of AiD+ and took notes on any sort of (technical) support provided.

n **Usability questionnaire filled out by GPs**: 4 technical usability criteria for the CDSS (learnability, clarity, handling, practicality in everyday use)[58] were rated on a 6-point Likert scale (ratings from 1=best to 6=worst) after the case vignette.

(9) **Individual semi-structured interviews with HCAs**: HCAs described their experience with the intervention; questions focused on the feasibility of the intervention, its components and the effects of the intervention on the practice team.

(10) **Intervention case vignettes with HCAs**:[57] After completion of the semi-structured interview, HCAs were presented with a case example of a hypothetical patient including: (1) written information (date of birth, height, weight, serum potassium, sodium and creatinine, health insurance number, a list of 8 drugs practice medication); (2) a collection of pill containers for the brown bag review. The HCA entered the data into AiD+. The interviewer checked the correct use of AiD+ and took notes on any sort of (technical) support provided.

o **Usability questionnaire filled out by HCAs**: four technical usability criteria for the CDSS (learnability, clarity, handling, practicality in everyday use)[58] were rated on a 6-point Likert scale (ratings from 1=best to 6=worst) after completion of the case vignette.

Figure 1 PaT plot of the PRIMUM pilot trial. GPs, general practitioners; HCA, healthcare assistant.

comparison between the intervention and control groups was made on the basis of the difference between MAI scores at baseline (T0) and 6 weeks after the beginning of the intervention (T1). Descriptive statistics and ICCs are provided for the baseline characteristics of practices and patients, as well as for the primary and secondary outcome measures. To analyse the differences between the intervention and control groups, linear mixed models were used. The results are presented as adjusted (for clustering) mean differences between groups with 95% CIs and p values, and the corresponding ICCs. Since this was a pilot study, the analysis of all result parameters remained primarily descriptive.

To determine the reliability of the MAI, the individual ratings were dichotomised into two groups, 'appropriate' versus 'inappropriate', in accordance with earlier suggestions: (1) the ratings '1' and '2' were considered to be 'appropriate' and '3' 'inappropriate',[62 76] (2) prescriptions rated as '1' were considered 'appropriate' and those rated as '2' or '3' 'inappropriate'.[71] Observer agreement and chance-adjusted agreement were calculated using κ-statistics, and alternative measures, such as the B-statistic and prevalence-adjusted bias-adjusted κ (PABAK), when the prevalence of positive ratings was low.[77–79]

RESULTS

Feasibility of the study

Recruitment and maintenance

Of the 692 potentially eligible patients from 20 general practices, 230 were selected at random and 100 were included (flow chart: online supplementary web–appendix 3). Of the 130 patients not included in the study, 67 were not invited because the recruitment target had already been reached, 41 did not meet the inclusion criteria, 20 refused to participate and 2 gave no reasons. In the intervention group, one patient at T1 (hospitalised) and four patients at T2 (three were hospitalised, one refused further participation) were lost to follow-up. In the control group, we lost two patients at T1 and subsequently at T2 (one died, one switched GP).

Study population

The GPs were mostly male (75%), had a median age of 57 years (range: 40–62 years) and were clinically experienced (on average 22 years). The median age of the HCAs was 42 years (20–58 years), and of the patients 75 years (64–93 years). The baseline characteristics of the study population are shown in table 1.

Outcome measures

At baseline, the outcome measures were balanced in both groups (table 2). Medication appropriateness: The vast majority of MAI ratings was 'appropriate' (see online supplementary web–appendix 4) and changes in mean MAI scores were small in both groups (table 3). Based on B-statistics, the intra-rater reliability for the MAI items ranged from 0.90 to 0.99, and inter-rater reliability from 0.83 to 0.94 (see online supplementary web–appendix 4). Mean differences in secondary outcomes between groups were small with wide two-sided 95% CIs. There was also no consistent trend across measures (table 4). Completeness of data: Outcome measures based on data from case report forms and patient telephone interviews including the MAI could be analysed almost completely. The proportion of missing values in secondary outcomes based on data from patient questionnaires ranged from 5% to 10% at baseline, 6% to 11% at T1 and 10% to 14% at T2. The VRS had the highest number of missing values (table 4).

Feasibility of the intervention

Perspective of GPs

In short questionnaires (figure 1, icon l), GPs reported a median time requirement of 35 min (IQR: 25—60′) per patient and that they were very satisfied or satisfied with 39/49 (80%) interventions, rather satisfied with 7/49 interventions (14%) and rather dissatisfied with 1/49 interventions (2%). Two interventions were not assessed. In semistructured interviews, 10 GPs (figure 1, icon 7) described the intervention as feasible or feasible with limitations: 9/10 reported positive experiences using the CDSS ('it is clearly structured, it is well-arranged'; 'I liked … the weightings (for alerts)'), 1/10 did not ('I did not feel comfortable with this programme…because I did not completely understand it'.). Five of 10 GPs reported that the GP–patient consultation was a positive experience ('clearly more systematic than regular consultations'; 'more often focused on adverse effects'; 'cooperation with patients has been improved') and 9/10 GPs experienced improved communication with HCAs ('I certainly talked more with the HCA about one or the other patient … because she wanted to give her feedback').

With the case vignettes (figure 1, icon 8), 7/10 GPs needed support in using the CDSS (support with a specific command: 5/7, major support: 2/7). To optimise medication for the case vignette, GPs used on average two of the four available CDSS alert functions (figure 1, icon 4). The number of prescriptions fell by 58%, potentially severe drug–drug interactions by 86% and inappropriate renal dosage adjustments by 71%. Inappropriate non-steroidal anti-inflammatory drugs prescriptions for the case vignette were stopped by 6/10 GPs and substituted with appropriate analgesics by 3/10 GPs. The technical usability of the CDSS (figure 1, icon n) was rated by GPs in median with 'good' for learnability (IQR: 1.25–2), clarity (1–2) and handling (2–2.75). The technical usability of the CDSS in everyday practice was assessed in median 4.5 (IQR 2.25–5) and GPs reported in interviews that the 'poor' rating was mainly due to a lack of connectivity with their practice software systems and the amount of time required.

Perspective of HCAs

In short questionnaires (figure 1, icon m), HCAs reported a median time requirement of 45 min (IQR: 33

Table 1 Baseline characteristics of practices and patients

Practices	Intervention group n=10	Control group n=10
Location: number (%)		
Urban	4 (40)	2 (20)
Suburban	5 (50)	4 (40)
Rural	1 (10)	4 (40)
Single-handed practices: number (%)	6 (60)	4 (40)
Panel size: number (%)		
Fewer than 1000	4 (40)	1 (10)
1000–1499	6 (60)	5 (50)
1500 or more	0	4 (40)
General practitioners		
Age (mean, SD)	54.8±7.41	54.25±5.31
Male sex: umber (%)	7 (70)	8 (80)
Board certificate: number (%)	10 (100)	9 (90)
General practitioner/general internist	7 (70)/3 (30)	7 (70)/2 (20)
Years of clinical experience (mean, SD)	22.6±11.44	20.6±8.55
Years at practice site (mean, SD)	19.3±8.83	18.4±10.20
Healthcare assistants		
Age (mean, SD)	40.0±11.81	39.2±13.64
Female sex: number (%)	10 (100)	10 (100)
Years of professional experience	18.8±12.2	16.3±10.56
Years at practice site	12.0±9.61	9.6±9.84
Full-time employment: number (%)	5 (50)	4 (40)
Patients	**n=50**	**n=50**
Age (mean, SD)	75.8±6.70	75.2±5.88
Female sex: number (%)	28 (56)	24 (48)
Covered by statutory health insurance: number (%)	46 (92)	50 (100)
Living with spouse or family: number (%)	32 (64)	35 (70)
Fending for themselves: number (%)	46 (94)	45 (92)
Home care situation good or very good (GP assessment): number (%)	44 (88)	44 (90)
CIRS sum score (mean, SD)	10.6±4.38	9.4±4.20
CIRS number of affected organ systems (mean, SD)	6.0±2.38	5.7±2.37
Number of chronic diseases* (mean, SD)	8.4±2.52	7.0±2.62
Charlson comorbidity score	4.5±2.64	4.5±2.46
Most common chronic diseases†: number (%)		
Hypertension including end organ affection	45 (90)	41 (82)
Diabetes mellitus	27 (54)	32 (64)
Lipid metabolism disorders	25 (50)	27 (54)
Chronic ischaemic heart disease	25 (50)	21 (42)
Osteoarthritis, joint arthrosis	17 (34)	18 (36)
Atherosclerosis/peripheral arterial disease	21 (42)	14 (28)
Asthma/chronic obstructive pulmonary disease	16 (32)	16 (32)
Cardiac arrhythmias	20 (40)	10 (20)
Gout and hyperuricaemia	23 (46)	6 (12)
Thyroid dysfunction	16 (32)	11 (22)
Chronic heart failure	11 (22)	13 (26)
Chronic gastro-oesophageal disease	13 (26)	11 (22)
Vision problems	11 (22)	12 (24)
Mental illnesses	9 (18)	13 (26)
Liver diseases	13 (26)	8 (16)
Falls: number of patients (%)	7 (14)	6 (12)
Previous hospitalisation: number of patients (%)	13 (26)	12 (24)
Number of previous days in hospital (mean, SD)	2.6±6.05	3.1±8.78
Potential ADR symptoms‡: number (%)		
Bleeding diathesis§	18 (37)	15 (30)
Ankle oedema	15 (31)	20 (40)

Continued

Table 1 Continued

Practices	Intervention group n=10	Control group n=10
Dizziness§	13 (27)	11 (22)
Dyspnoea§	8 (16)	12 (24)
Difficulties urinating	6 (12)	10 (20)
Abdominal pain§	5 (10)	4 (8)
Tachycardia or palpitation§	4 (8)	4 (8)
Nausea or vomiting§	4 (8)	3 (6)

*Disease count according to the list of chronic diseases by Schafer *et al*.[87]
†Chronic diseases[87] prevalent in more than 20% of the total study population in descending order.
‡Symptoms appeared on several days or almost every day.
§For details, see figure 1.
ADR, adverse drug reaction; CIRS, Cumulative Illness Rating Scale; GP, general practitioner.

—70′) and were very satisfied or satisfied in 92% of cases (45/49), and rather satisfied in 2/49 cases (4%). No intervention was considered rather dissatisfying or worse, and two interventions were not assessed. In semi-structured interviews, HCAs (figure 1, icon 9) reported no major problems with the intervention and positive experiences with the patients: 9/10 HCAs had no difficulties using and filling out the MediMoL ('I really had no problems, it all went well'), one had difficulties ('Not all the questions were clear to me'). The CDSS performed well: 9/10 HCAs described the experience as 'very good' ('I could use it very easily, I am doing fine with it'), one considered the experience 'rather good' ('It would be nice, if (the CDSS) would transfer (the medication) …from Medibox 1 to Medibox 2'). The HCAs felt the investigator and intervention trainings (figure 1, icons 1 and 2) prepared them well for the study ('The tasks were clearly described and well-

Table 2 Outcomes at baseline

	Intervention group (10 practices, 50 patients)		Control group (10 practices, 50 patients)		
	n_i	M_i (SD)	n_c	M_c (SD)	ICC/ICC_{adj}
MAI	50	5.0 (4.69)	50	4.8 (5.40)	0.085/0.099
EQ-5D, index	45	0.80 (0.19)	48	0.80 (0.18)	0.000/0.006
VRS	43	4.8 (2.37)	47	4.4 (2.67)	0.021/0.029
GDS	49	2.2 (2.12)	49	1.7 (1.89)	0.029/0.025
Number of drugs	50	9.5 (2.67)	50	8.7 (2.66)	0.186/0.180
Number of single doses	50	9.5 (3.19)	50	9.4 (3.83)	0.177 0.192
MRCI	50	13.8 (7.21)	50	13.7 (5.98)	0.048/0.060
Number of ADR symptoms	49	1.9 (1.62)	50	1.8 (1.44)	0.031/0.040
Observed adherence*					
Drug score (%)†	48	20 (41.7)	49	23 (46.9)	0.003/0.001
Dose score (%)†	46	45 (97.8)	48	48 (100)	–/–‡
Regimen score (%)†	45	45 (100)	48	48 (100)	–/–‡
Reported adherence					
MARS	46	23.9 (1.68)	47	24 (1.16)	0.0004/0.008
Morisky	49	0.3 (0.69)	50	0.2 (0.48)	0.000/0.000
BMQ					
Specific necessities	45	22.4 (3.26)	48	22.5 (2.57)	0.000/0.000
Specific concerns	44	14 (5.52)	47	13 (4.91)	0.000/0.000
General overuse	44	10.3 (4.77)	47	10.4 (3.88)	0.000/0.006
General harms	47	7.8 (3.40)	48	7.6 (3.30)	0.000/0.000

Numbers of patients (n), mean (M), SD and intracluster correlation coefficients (ICCs) are provided, the latter as crude values and adjusted for groups.
*Discrepancy between medication actually taken (reported at patient's interviews) and medication prescribed (reported by GP).
†Number and percentage of deviating patients.
‡No ICC could be estimated, as (almost) all patients were scored as deviating.
ADR, adverse drug reaction; BMQ, Beliefs in Medicine Questionnaire; CIRS, Cumulative Illness Rating Scale; GDS, Geriatric Depression Scale; MAI, Medication Appropriateness Index; MARS, Medication Adherence Rating Scale; MRCI, Medication Regimen Complexity Index; VRS, Visual Rating Scale (pain assessment).

Table 3 Outcome MAI

MAI sum score at patient level	Intervention group n_i	M_i (SD)	$Median_i$ (IQR)	Control group n_c	M_c (SD)	$Median_c$ (IQR)	ICC/ICC_{adj}	p Value
T0	50	5.0 (4.69)	3.5 (2–6.75)	50	4.8 (5.40)	3 (1–6.75)	0.085/0.099	0.882
T1	48	5.1 (5.12)	4 (1–7.25)	48	4.2 (5.09)	2 (0–7)	0.000/0.000	0.401
T2	46	5.7 (6.57)	3 (1–7)	47	4.6 (4.76)	3 (1–7)	0.049/0.053	0.387
MAI sum score T1-T0	48	0.1 (5.31)	0.5 (–3–3)	48	–0.6 (5.56)	0 (–2.25–1.25)	0.023/0.032	0.548
M_I-M_C (95% CI), not adjusted						–0.7 (–1.6 to 0.2)*		
M_I-M_C (95% CI), adjusted for clustering						–0.7 (–3.0 to 1.6)*		
MAI sum score T2-T0	46	0.7 (5.45)	0 (–2–3)	47	–0.2 (5.17)	0 (–2–2.5)	0.030/0.039	0.460
M_I-M_C (95% CI), not adjusted						–0.9 (–1.8 to 0.1)*		
M_I-M_C (95% CI), adjusted for clustering						–0.9 (–3.2 to 1.4)*		
Inappropriate prescriptions per group		**Number n_i/N_i**	**Proportion %**		**Number n_c/N_c**	**Proportions %**		
T0		130/439	29.6		108/412	26.2		
T1		109/415	26.3		90/385	23.4		
T2		107/393	27.2		99/371	26.7		
Inappropriate prescriptions per patient		**Number M_i (SD)**	**Proportion M_i (SD)**		**Number M_c (SD)**	**Proportion M_c (SD)**		
T0		2.60 (1.67)	0.30 (0.18)		2.16 (1.80)	0.25 (0.19)		
T1		2.27 (1.82)	0.26 (0.20)		1.875 (1.81)	0.22 (0.20)		
T2		2.33 (1.89)	0.26 (0.20)		2.11 (1.82)	0.24 (0.20)		

If not stated otherwise, mean (M) and SD are provided. ICCs are provided as crude values and adjusted for group. p Values are adjusted for cluster effects.
*The trend was in favour of the control group.
ICC, intracluster correlation coefficient; MAI, medication appropriateness index.

Table 4 Secondary outcomes

	T1-T0							T2-T0						
	Intervention group		Control group					Intervention group		Control group				
	n_i	M_i (SD)	n_c	M_c (SD)	ICC/ICC_{adj}	p Value	M_i—M_c (95%-CI)	n_i	M_i (SD)	n_c	M_c (SD)	ICC/ICC_{adj}	p Value	M_i_M_c (95%-CI)
EQ-5D, index	40	0.0 (15.89)	42	2.4 (17.85)	0.022/.037	0.531	2.5 (−5.2 to 10.3)	40	−0.6 (19.61)	36	−1 (13.66)	0.000/0.000	0.926	−0.4 (−8 to 7.3)
VRS	36	0.0 (1.71)	42	0.0 (2.32)	0.000/0.008	0.968	0.0 (−0.9 to 0.9)	36	0.0 (1.84)	37	0.0 (2.69)	0.000/0.008	0.968	0.0 (−0.9 to 0.9)
GDS	46	−0.1 (1.84)	46	0.4 (1.71)	0.000/0.000	0.193	0.5 (−0.2 to 1.2)	43	−0.4 (1.92)	46	1.2 (2.28)	0.201/0.098	0.006	1.6 (0.6 to 2.7)
Number of drugs	49	−0.1 (2.41)	48	−0.4 (1.39)	0.000/.000	0.569	−0.2 (−1.0 to 0.6)	46	−0.7 (1.56)	47	−0.4 (1.43)	0.063/0.069	0.497	0.2 (−0.4 to 0.9)
Number of single doses	49	−0.1 (2.21)	48	−0.3 (2.8)	0.140/0.153	0.753	−0.2 (−1.5 to 1.1)	46	0.2 (1.79)	47	−0.4 (2.65)	0.060/0.062	0.302	−0.6 (−1.6 to 0.5)
MRCI	49	1.3 (5.43)	48	0.3 (3.26)	0.079/0.078	0.374	−1.0 (−3.0 to 1.1)	46	1.4 (5.24)	47	0.6 (4.91)	0.091/0.100	0.555	−0.7 (−3.2 to 1.7)
Number of ADR symptoms	46	0.1 (1.48)	47	0.2 (1.30)	0.131/0.147	0.844	−0.1 (−0.7 to 0.8)	43	0.0 (1.02)	47	−0.2 (1.34)	0.084/0.093	0.569	−0.2 (10.8 to 0.4)
Observed adherence*:														
Drug score (%)	45	0.0 (0.35)	47	0.1 (0.28)	0.054/.065	0.560	0.0 (−0.1 to 0.2)	42	0.0 (0.33)	46	0.0 (0.28)	0.128/0.142	0.636	0.0 (−0.1 to 0.2)
Dose score (%)	42	0.0 (0.23)	45	0.0 (0.21)	0.042/.051	0.570	0.0 (−0.1 to 0.1)	39	0.0 (0.23)	45	0.0 (0.21)	0.033/0.047	0.959	0.0 (−0.1 to 0.1)
Regimen score (%)	41	0.0 (0.14)	45	0.0 (0.18)	0.000/.000	0.977	0.0 (−0.1 to 0.1)	38	0.0 (0.14)	45	0.0 (0.19)	0.000/0.000	0.761	0.0 (−0.1 to 0.1)
Reported adherence:														
MARS	42	0.0 (1.64)	43	−0.1 (1.18)	0.000/0.000	0.944	0.0 (−0.6 to 0.6)	43	−0.2 (1.79)	37	0.1 (1.14)	0.000/0.000	0.484	0.2 (−0.4 to 0.9)
Morisky	46	−0.1 (0.64)	47	0.0 (0.55)	0.000/0.000	0.237	0.2 (−0.1 to 0.4)	43	0.0 (0.79)	47	0.0 (0.47)	0.000/0.000	1.00	0.0 (−0.3 to 0.3)
BMQ:														
Specific necessities	42	−0.4 (2.84)	43	−0.7 (3.55)	0.047/0.056	0.693	−0.3 (−1.8 to 1.2)	40	0.1 (3.59)	38	−1.0 (4.72)	0.231/0.239	0.277	−1.4 (−3.9 to 1.1)
Specific concerns	40	−1.2 (5.03)	42	−0.9 (5.35)	0.185/.200	0.934	0.1 (−2.8 to 3.0)	39	−1.3 (6.17)	36	−1.6 (5.43)	0.275/0.294	0.724	−0.7 (−4.3 to 3)
General overuse	39	0.1 (3.18)	40	−0.1 (2.76)	0.000/.000	0.766	−0.2 (−1.5 to 1.1)	39	0.2 (4.18)	37	−0.8 (2.42)	0.000/0.000	0.251	−0.1 (−1.9 to 1.7)
General harms	44	0 (2.26)	41	−0.4 (2.85)	0.000/0.000	0.395	−0.5 (−1.6 to 0.6)	42	−0.2 (3.45)	36	−0.5 (3.17)	0.000/0.012	0.664	0.2 (−1.2 to 1.5)
Number of hospital stays†	50	−0.4 (0.73)	50	−0.2 (0.62)				–	–	–	–	–	–	–
Days in hospital†	50	−0.4 (0.73)	50	−0.2 (0.69)				–	–	–	–	–	–	–

If not stated otherwise, mean and SD are provided. ICCs are provided as crude values and adjusted for group. p Values are adjusted for cluster effects.
*Discrepancy between medication actually taken (reported at patient's interviews) and medication prescribed (reported by GP).
†Hospitalisations are aggregated for both follow-ups: T1+T2−T0; due to low event rates and skewness of distribution; no ICCs, p values and mean differences are provided.
ADR, adverse drug reaction; BMQ, Beliefs in Medicine Questionnaire; CIRS, Cumulative Illness Rating Scale; GDS, Geriatric Depression Scale; ICCs, intracluster correlation coefficients; MAI, Medication Appropriateness Index; MARS, Medication Adherence Reporting Scale; MRCI, Medication Regimen Complexity Index; VRS, Visual Rating Scale (pain assessment).

structured; I had no problems'). The encounter with the patient was assessed particularly positively ('I really liked being allowed to do the tests on the patients and being able to work so closely with them').

The case vignette (figure 1, icon 10) was understood by 6/10 HCAs without any help, while 4/10 HCAs needed minor support (eg, to enter the name of a complementary drug formulation in Medibox 2). The technical usability of the CDSS (figure 1, icon o) was rated by the 10 HCAs in median with 'good' for all dimensions, with the IQR regarded as slightly better for learnability and handling (1.25–2) than for clarity and workaday practicability (2–2).

Patients' perspective

In telephone interviews, 23/42 patients knew what the study was about and explained the potential benefits for themselves ('Drug tolerance and interactions between different drugs are looked at, to see if there might be one that could be left out'), whereas the remaining patients did not understand the study and did not feel they had benefited from participating in it ('I'm not reckoning on benefiting from it personally, but my doctor asked me to'). None of the patients said that any of the questions asked by the HCA in the preconsultation interview had made them feel uneasy. Patients' symptoms were mostly about the study methods ('Being asked twice about medicines; only one of mine was dropped; that could have been decided quicker'; 'It's difficult to answer all the questions in the questionnaire with a simple yes or no').

DISCUSSION

The complex intervention to prioritise and optimise multimedication in older patients with multimorbidity, and the study design, are feasible in a general practice setting. However, the pilot study revealed a number of limitations and potential barriers to the future implementation of the complex intervention that should be addressed when designing the main trial.

Feasibility of the complex intervention

Participating GPs valued the structured systematic approach to conducting consultations and said working relationships with patients and HCAs had improved. HCAs appreciated being involved in the complex intervention. Both GPs and HCAs reported mainly positive experiences with the tools MediMoL and the CDSS and rated the (technical) usability of the CDSS as 'good'. Moreover, GPs were often surprised by the discrepancy between prescribed and taken medicines, as confirmed by the brown bag review. Only slightly more than half the study patients were fully aware of the rationale and the aims of the study. However, informed patients welcomed the chance to detect inappropriate prescriptions and to adjust their medication. Nevertheless, GPs pointed out that the process required considerable time and said the incompatibility of the CDSS with their practice software was a relevant barrier to future practice implementation. Positive results in interviews and questionnaires differed somewhat from the situation with case vignettes where difficulties were experienced using the CDSS application: more GPs than HCAs needed help in using the features and running the programme. Most HCAs and GPs did not use the CDSS following the completion of the final intervention, so their difficulties may have resulted from a lack of training and the time lag between the final intervention and the case vignette (figure 1). Since we did not provide a manual, it is possible that not all practices correctly implemented the CDSS.

Feasibility of the trial design

Most procedures went well—recruitment was completed with equal cluster sizes, randomisation resulted in overall balanced groups, loss to follow-up was within acceptable limits and data collection and the medication reviews by the clinical pharmacologist were feasible. Missing data were most common in patients' questionnaires, and in the VRS in particular. Patients' interviews showed that some patients had difficulties understanding questions from the validated instruments. The most relevant outcome measures, MAI and EQ-5D, showed an almost perfect baseline value, leaving little room for improvement. First, cardiovascular comorbidity was highly prevalent, with common diseases sharing the same pathways and treatment targets. This may have prevented GPs from having to deal with potentially harmful interactions. Second, a reduction in inappropriate prescriptions was observed in both groups, indicating a likely contamination effect in the control group: both groups received the study protocol including a detailed description of the intervention. Although the CDSS was only available to the intervention group, the control group may have conducted brown bag reviews and medication reviews with or without computer support. Some practice software provides alert features for drug–drug interactions. However, these are often deactivated due to over-alerting.[80]

Furthermore, the low prevalence of 'inappropriate prescriptions' and imbalanced marginals for MAI ratings in our sample led to paradoxically low κ values despite high intra-rater and inter-rater observer agreement.[79] In this situation, alternative reliability measures such as B-statistics and PABAK are recommended.[77–79] Using these measures, intra-rater reliability of MAI ratings showed almost perfect agreement[81] and intra-rater reliability was slightly better, which is in line with former observations.[62 69 73] Evaluated secondary outcomes showed small changes but supported for most of them a further use in the main study (EQ-5D and adherence-related measures such as medication complexity). As observed in earlier studies,[82] measures of self-reported adherence did not appear to provide valid results, as they contradicted results from comparisons of

prescribed with taken medicines, and showed ceiling versus floor effects. Additionally, MARS had a large number of missing values. The functionality outcome (WHO-DAS II) was not usable, because the manual was under development and did not provide a correct formula.

The application of the cluster-RCT design was both a strength, because it allowed to put in place all procedures of the planned main study, and a challenge, because the integration of a control group doubled the sample size for feasibility testing of the complex intervention. Furthermore, participants may have overestimated the time required by the intervention because the time required for data collection and other procedures may have been included in estimates of the time required to perform the complex intervention. With the use of mixed methods, however, we were able to identify obstacles to the complex intervention and its implementation that helped us to improve the design of the main study. The CDSS recorded data on use (eg, date of use, completion of Mediboxes by GPs/HCAs), but these data did not provide information on whether or not the users correctly applied the different features to check for interactions, appropriate dosage, etc. In qualitative interviews, GPs and HCAs did not report problems using the CDSS when asked. However, case vignettes helped detect difficulties experienced by GPs and HCAs in the use of the new software. These can be eliminated by intensifying training and providing supporting material. Limited resources prevented us from gaining detailed insights into usual care provided by GPs when adjusting medication for older multimorbid patients, and this is a further limitation of our study. This information could have been helpful in planning the main trial.

This article provides the results of the systematic piloting of a complex intervention for polypharmacy and its corresponding trial design in primary care. Published trials on complex interventions in polypharmacy included in a current Cochrane review were not piloted at all or mentioned only a piloting phase without describing results and conclusions.[12] Many of the studies were conducted after publication of the MRC guidance that strongly recommends a piloting phase.[30 83] Very recently, Clyne and coauthors reported on an alternative approach, also aimed at helping in the development of a complex intervention to reduce potentially inappropriate prescribing (PIP), but which uses explicit criteria.[84] The authors described an exhaustive consensus process for deriving an acceptable set of PIPs from lists identified in a literature review (eg, Beers and STOPP criteria[85 86]). Focusing on a high acceptability at the provider level, the authors applied predominantly qualitative methods that resulted in the stepwise improvement of the intervention. Our approaches differed mainly in purpose, methods and (presumably) in cost but both highlight the fact that descriptions of piloting phases are particularly useful for a number of reasons: they typically use more diverse techniques than full studies, uncover critical pitfalls and challenges and provide important insights into promising techniques, facilitators and barriers and often also into the causes of success and failure.

Lessons learnt

Feasibility testing of our complex intervention has enabled us to improve the design of the main study: as a consequence, investigator training has been intensified and supported by a written manual with a strong focus on using the CDSS. The multitude of used interfaces will prevent significant improvement in connectivity to practice software systems in the main study. The National Association of Statutory Health Insurance Physicians has only recently begun to work on harmonising data interfaces for manufacturers of practice software, thus facilitating future data exchange with systems such as our CDSS.

Feasibility testing identified a potential contamination problem with the control group. We have therefore decided that in the main study, no details of the intervention will be shared with healthcare professionals of the control group. Furthermore, we have changed the inclusion criteria. To include a greater number of patients at risk of (manageable) interactions, patients have to have not only three or more chronic diseases, but the diseases must be from at least two different chapters of ICD-10. We have also replaced impractical outcome measures (VRS, WHO-DAS II and MARS).

Although we have demonstrated feasibility and potential limitations of the complex intervention, its effectiveness in general practice has yet to be proven. Furthermore, it is as yet unclear whether the advantages will outweigh the disadvantages in terms of required time and costs, and whether the barriers to a wider implementation in routine care can be removed.

CONCLUSION

Our pilot study of a complex intervention to prioritise and optimise multimedication in older patients with multimorbidity has confirmed the feasibility of the intervention and the study design, but has also revealed rather important limitations and options for improvement. These have enabled us to refine and modify the final design and improve the main study in critical areas such as measures to limit contamination, inclusion criteria and outcome measures.

Author affiliations
[1]Institute of General Practice, Johann Wolfgang Goethe University, Frankfurt/Main, Germany
[2]Institute for Clinical Pharmacology, Johann Wolfgang Goethe University Hospital, Frankfurt/Main, Germany
[3]Institute of Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany
[4]Department of Family Medicine, School CAPHRI, Maastricht University, Maastricht, The Netherlands
[5]Department of General Practice, KU Leuven, Leuven, Belgium
[6]Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

[7]Health Services & Policy Research Group, School of Medicine, University of Exeter, Exeter, UK
[8]Department of Clinical Pharmacology and Pharmacoepidemiology, University of Heidelberg, Heidelberg, Germany

Acknowledgements The authors would like to thank all participating patients and all general practice investigators and their teams (Drs Bauer, Bolender, Braun, Dörr, Draheim, Endruweit, Fink, Freise, Gerlach-Lüdeke, Göllner, Heiskel, Jablonski, Neuschild, Roser, Rothkegel, Sanner Schiek-Kunz, Sunnus, Thürmer, Vetter, Weismüller) and Petra Thuermann for conducting medication reviews during the MAI reliability study. The authors are grateful to our dedicated research team Zeycan Albay, Anja Paesel, Mareike Leifermann, Anne Namyst and to Gisela Kassner. The authors also thank our practice advisory board Joachim Fessler, Joachim Seffrin, Karola Mergenthal and Vera Müller. This study was only possible thanks to their extraordinary engagement. Furthermore, the authors would like to acknowledge the tremendous work of Kristina Zint, Jens Kaltschmidt and Michael Metzner in developing and programming the study version of the CDSS. The authors thank Sven Schulz and Jochen Gensichen for the prepiloting of the MediMoL and data collection instruments and Cornelia Mahler for the provision of the German versions of MARS and BMQ. The authors would also like to thank Professor Paul Glasziou and the participants of the journal club at CREBP, Bond University, QLD, Australia for inspiring discussions on the MAI and particularly Elaine Beller for her support in the sample size calculation of the MAI reliability study. Further, the authors would like to acknowledge the work of Phillip Elliott in translating the MediMoL and quotations from interviews, and proofreading the manuscript.

Contributors CM drafted the manuscript, coordinated the study and contributed to the conception, design, data collection and data analyses. JR contributed to the conception, design and data analyses. WEH contributed to the conception and design and provided the study version of CDSS. SH contributed to the conception and design and conducted the MAI ratings. BF and LU contributed to the design, data collection and analyses. CG, AE and MB contributed to the conception and design and supported the recruitment of practices. RP, MvdA, AK, JVM and FMG provided advice on the conception, design and coordination of the study. All authors critically revised and agreed on the final version of the manuscript.

Funding Funding has been provided by the German Federal Ministry of Education and Research, BMBF, grant number 01GK0702.

Competing interests CM, LU, BF, CMV, AE, CG, JAK, JR, RP, SH and MB have nothing to disclose. WEH is a member of the scientific advisory board and a shareholder of Dosing GmbH, the company distributing the clinical decision support software used in this study. His wife is an employee of Dosing GmbH.

Ethics approval Institutional Review Board at University Hospital, Goethe University, number 54/09, 24/03/2009.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Salisbury C, Johnson L, Purdy S, *et al*. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *Br J Gen Pract* 2011;61:e12–21.
2. Kuijpers MA, van Marum RJ, Egberts AC, *et al*. Relationship between polypharmacy and underprescribing. *Br J Clin Pharmacol* 2008;65:130–3.
3. Nobili A, Marengoni A, Tettamanti M, *et al*. Association between clusters of diseases and polypharmacy in hospitalized elderly patients: results from the REPOSI study. *Eur J Intern Med* 2011;22:597–602.
4. Campbell SE, Seymour DG, Primrose WR, *et al*. A multi-centre European study of factors affecting the discharge destination of older people admitted to hospital: analysis of in-hospital data from the ACMEplus project. *Age Ageing* 2005;34:467–75.
5. Hanlon JT, Schmader KE, Koronkowski MJ, *et al*. Adverse drug events in high risk older outpatients. *J Am Geriatr Soc* 1997;45:945–8.
6. Schneeweiss S, Hasford J, Göttler M, *et al*. Admissions caused by adverse drug events to internal medicine and emergency departments in hospitals: a longitudinal population-based study. *Eur J Clin Pharmacol* 2002;58:285–91.
7. Thomsen LA, Winterstein AG, Søndergaard B, *et al*. Systematic review of the incidence and characteristics of preventable adverse drug events in ambulatory care. *Ann Pharmacother* 2007;41:1411–26.
8. Col N, Fanale JE, Kronholm P. The role of medication noncompliance and adverse drug reactions in hospitalizations of the elderly. *Arch Intern Med* 1990;150:841–5.
9. Gurwitz JH, Field TS, Harrold LR, *et al*. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *JAMA* 2003;289:1107–16.
10. Davies EC, Green CF, Taylor S, *et al*. Adverse drug reactions in hospital inpatients: a prospective analysis of 3695 patient-episodes. *PLoS ONE* 2009;4:e4439.
11. Pirmohamed M, James S, Meakin S, *et al*. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* 2004;329:15–19.
12. Patterson SM, Cadogan CA, Kerse N, *et al*. Interventions to improve the appropriate use of polypharmacy for older people. *Cochrane Database Syst Rev* 2014;(10):CD008165.
13. Smith SM, Wallace E, O'Dowd T, *et al*. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. *Cochrane Database Syst Rev* 2016;3:CD006560.
14. Bergman-Evans B. Evidence-based guideline. Improving medication management for older adult clients. *J Gerontol Nurs* 2006;32:6–14.
15. Muth C, Beyer M, Zint K, *et al*. A matter of attention: inconsistencies between prescription and drug intake in elderly multimorbid patients in primary care. *Eur J General Practice* 2009;15:46.
16. Schmiemann G, Bahr M, Gurjanov A, *et al*. Differences between patient medication records held by general practitioners and the drugs actually consumed by the patients. *Int J Clin Pharmacol Ther* 2012;50:614–17.
17. Fillit HM, Futterman R, Orland BI, *et al*. Polypharmacy management in Medicare managed care: changes in prescribing by primary care physicians resulting from a program promoting medication reviews. *Am J Manag Care* 1999;5:587–94.
18. Garcia RM. Five ways you can reduce inappropriate prescribing in the elderly: a systematic review. *J Fam Pract* 2006;55:305–12.
19. Patterson SM, Hughes C, Kerse N, *et al*. Interventions to improve the appropriate use of polypharmacy for older people. *Cochrane Database Syst Rev* 2012;(5):CD008165.
20. Bryan C, Boren SA. The use and effectiveness of electronic clinical decision support tools in the ambulatory/primary care setting: a systematic review of the literature. *Inform Prim Care* 2008;16:79–91.
21. Garg AX, Adhikari NK, McDonald H, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293:1223–38.
22. Pearson SA, Moxey A, Robertson J, *et al*. Do computerised clinical decision support systems for prescribing change practice? A systematic review of the literature (1990–2007). *BMC Health Serv Res* 2009;9:154.
23. Yourman L, Concato J, Agostini JV. Use of computer decision support interventions to improve medication prescribing in older adults: a systematic review. *Am J Geriatr Pharmacother* 2008;6:119–29.
24. Wetzels R, Harmsen M, Van Weel WC, *et al*. Interventions for improving older patients' involvement in primary care episodes. *Cochrane Database Syst Rev* 2007;(1):CD004273.
25. Bosley S, Dale J. Healthcare assistants in general practice: practical and conceptual issues of skill-mix change. *Br J Gen Pract* 2008;58:118–24.
26. Gensichen J, Jaeger C, Peitz M, *et al*. Health care assistants in primary care depression management: role perception, burdening factors, and disease conception. *Ann Fam Med* 2009;7:513–19.
27. Gensichen J, von Korff KM, Peitz M, *et al*, PRoMPT (PRimary care Monitoring for depressive Patients Trial). Case management for depression by health care assistants in small primary care practices: a cluster randomized trial. *Ann Intern Med* 2009;151:369–78.
28. Peters-Klimm F, Muller-Tasch T, Schellberg D, *et al*. Rationale, design and conduct of a randomised controlled trial evaluating a

primary care-based complex intervention to improve the quality of life of heart failure patients: HICMan (Heidelberg Integrated Case Management). *BMC Cardiovasc Disord* 2007;7:25.
29. Rosemann T, Körner T, Wensing M, *et al*. Rationale, design and conduct of a comprehensive evaluation of a primary care based intervention to improve the quality of life of osteoarthritis patients. The PraxArt-project: a cluster randomized controlled trial [ISRCTN87252339]. *BMC Public Health* 2005;5:77.
30. Craig P, Dieppe P, Macintyre S, *et al*. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;337:a1655.
31. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract* 2004;10:307–12.
32. Lancaster GA, Campbell MJ, Eldridge S, *et al*. Trials in primary care: statistical issues in the design, conduct and evaluation of complex interventions. *Stat Methods Med Res* 2010;19:349–77.
33. Hudon C, Fortin M, Soubhi H. Abbreviated guidelines for scoring the Cumulative Illness Rating Scale (CIRS) in family practice. *J Clin Epidemiol* 2007;60:212.
34. Lin H, Allore HG, McAvay G, *et al*. A method for partitioning the attributable fraction of multiple time-dependent coexisting risk factors for an adverse health outcome. *Am J Public Health* 2013;103:177–82.
35. Bergert FW, Conrad D, Ehrenthal K, *et al*. Pharmacotherapy guidelines for the aged by family doctors for the use of family doctors. *Int J Clin Pharmacol Ther* 2008;46:600–16.
36. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–98.
37. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
38. Chwastiak LA, Von Korff M. Disability in depression and back pain: evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting. *J Clin Epidemiol* 2003;56:507–14.
39. Janca A, Kastrup M, Katschnig H, *et al*. The World Health Organization Short Disability Assessment Schedule (WHO DAS-S): a tool for the assessment of difficulties in selected areas of functioning of patients with mental disorders. *Soc Psychiatry Psychiatr Epidemiol* 1996;31:349–54.
40. Horne R, Weinman J. Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. *J Psychosom Res* 1999;47:555–67.
41. Horne R, Weinman J, Hankins M. The Beliefs About Medicines Questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. *Psychol Health* 1999;14:1–24.
42. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care* 1986;24:67–74.
43. Sheikh JI, Yesavage JA, Brooks JO, III, *et al*. Proposed factor structure of the Geriatric Depression Scale. *Int Psychogeriatr* 1991;3:23–8.
44. Yesavage JA, Brink TL, Rose TL, *et al*. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1982;17:37–49.
45. Cascorbi I. Drug interactions—principles, examples and clinical consequences. *Dtsch Arztebl Int* 2012;109:546–55.
46. Juurlink DN, Mamdani M, Kopp A, *et al*. Drug-drug interactions among elderly patients hospitalized for drug toxicity. *JAMA* 2003;289:1652–8.
47. Wrenger E, Müller R, Moesenthin M, *et al*. Interaction of spironolactone with ACE inhibitors or angiotensin receptor blockers: analysis of 44 cases. *BMJ* 2003;327:147–9.
48. Harder S, Thürmann P. Clinically important drug interactions with anticoagulants. An update. *Clin Pharmacokinet* 1996;30:416–44.
49. Hines LE, Murphy JE. Potentially harmful drug-drug interactions in the elderly: a review. *Am J Geriatr Pharmacother* 2011;9:364–77.
50. Bergk V, Haefeli WE, Gasse C, *et al*. Information deficits in the summary of product characteristics preclude an optimal management of drug interactions: a comparison with evidence from the literature. *Eur J Clin Pharmacol* 2005;61:327–35.
51. Bertsche T, Fleischer M, Pfaff J, *et al*. Pro-active provision of drug information as a technique to address overdosing in intensive-care patients with renal insufficiency. *Eur J Clin Pharmacol* 2009;65:823–9.
52. Bertsche T, Pfaff J, Schiller P, *et al*. Prevention of adverse drug reactions in intensive care patients by personal intervention based on an electronic clinical decision support system. *Intensive Care Med* 2010;36:665–72.
53. Falconnier AD, Haefeli WE, Schoenenberger RA, *et al*. Drug dosage in patients with renal failure optimized by immediate concurrent feedback. *J Gen Intern Med* 2001;16:369–75.
54. Martin-Facklam M, Rengelshausen J, Tayrouz Y, *et al*. Dose individualisation in patients with renal insufficiency: does drug labelling support optimal management? *Eur J Clin Pharmacol* 2005;60:807–11.
55. Martin P, Haefeli WE, Martin-Facklam M. A drug database model as a central element for computer-supported dose adjustment within a CPOE system. *J Am Med Inform Assoc* 2004;11:427–32.
56. Quinzler R, Schmitt SP, Szecsenyi J, *et al*. Optimizing information on drug exposure by collection of package code information in questionnaire surveys. *Pharmacoepidemiol Drug Saf* 2007;16:1024–30.
57. Hoddinott P, Morgan H, Thomson G, *et al*. Intervention vignettes as a qualitative tool to refine complex intervention design. *Trials* 2013;14(Suppl 1):O55.
58. Daniels J, Fels S, Kushniruk A, *et al*. A framework for evaluating usability of clinical monitoring technology. *J Clin Monit Comput* 2007;21:323–30.
59. Perera R, Heneghan C, Yudkin P. Graphical method for depicting randomised trials of complex interventions. *BMJ* 2007;334:127–9.
60. Spinewine A, Schmader KE, Barber N, *et al*. Appropriate prescribing in elderly people: how well can it be measured and optimised? *Lancet* 2007;370:173–84.
61. Fitzgerald LS, Hanlon JT, Shelton PS, *et al*. Reliability of a modified medication appropriateness index in ambulatory older persons. *Ann Pharmacother* 1997;31:543–8.
62. Hanlon JT, Schmader KE, Samsa GP, *et al*. A method for assessing drug therapy appropriateness. *J Clin Epidemiol* 1992;45:1045–51.
63. Barat I, Andreasen F, Damsgaard EM. Drug therapy in the elderly: what doctors believe and patients actually do. *Br J Clin Pharmacol* 2001;51:615–22.
64. George J, Phun YT, Bailey MJ, *et al*. Development and validation of the medication regimen complexity index. *Ann Pharmacother* 2004;38:1369–76.
65. Neergaard MA, Olesen F, Andersen RS, *et al*. Qualitative description—the poor cousin of health research? *BMC Med Res Methodol* 2009;9:52.
66. Sandelowski M. What's in a name? Qualitative description revisited. *Res Nurs Health* 2010;33:77–84.
67. Thabane L, Ma J, Chu R, *et al*. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol* 2010;10:1.
68. Browne RH. On the use of a pilot sample for sample size determination. *Stat Med* 1995;14:1933–40.
69. Bregnhøj L, Thirstrup S, Kristensen MB, *et al*. Reliability of a modified medication appropriateness index in primary care. *Eur J Clin Pharmacol* 2005;61:769–73.
70. Gallagher PF, O'Connor MN, O'Mahony D. Prevention of potentially inappropriate prescribing for elderly patients: a randomized controlled trial using STOPP/START criteria. *Clin Pharmacol Ther* 2011;89:845–54.
71. Kassam R, Martin LG, Farris KB. Reliability of a modified medication appropriateness index in community pharmacies. *Ann Pharmacother* 2003;37:40–6.
72. Spinewine A, Dumont C, Mallet L, *et al*. Medication appropriateness index: reliability and recommendations for future use. *J Am Geriatr Soc* 2006;54:720–2.
73. Stuijt CC, Franssen EJ, Egberts AC, *et al*. Reliability of the medication appropriateness index in Dutch residential home. *Pharm World Sci* 2009;31:380–6.
74. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257–68.
75. Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation. *Stat Med* 1992;11:1511–19.
76. Samsa GP, Hanlon JT, Schmader KE, *et al*. A summated score for the medication appropriateness index: development and assessment of clinimetric properties including content validity. *J Clin Epidemiol* 1994;47:891–6.
77. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423–9.
78. Mak HK, Yau KK, Chan BP. Prevalence-adjusted bias-adjusted kappa values as additional indicators to measure observer agreement. *Radiology* 2004;232:302–3.
79. Shankar V, Bangdiwala SI. Observer agreement paradoxes in 2x2 tables: comparison of agreement measures. *BMC Med Res Methodol* 2014;14:100.

80. Seidling HM, Klein U, Schaier M, *et al*. What, if all alerts were specific—estimating the potential impact on drug interaction alert burden. *Int J Med Inform* 2014;83:285–91.
81. Munoz SR, Bangdiwala SI. Interpretation of Kappa- and B statistics measures of agreement. *J Appl Stat* 1997;24:105–11.
82. Jank S, Bertsche T, Schellberg D, *et al*. The A14-scale: development and evaluation of a questionnaire for assessment of adherence and individual barriers. *Pharm World Sci* 2009;31:426–31.
83. Campbell M, Fitzpatrick R, Haines A, *et al*. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694–6.
84. Clyne B, Bradley MC, Hughes CM, *et al*, OPTI-SCRIPT study team. Addressing potentially inappropriate prescribing in older patients: development and pilot study of an intervention in primary care (the OPTI-SCRIPT study). *BMC Health Serv Res* 2013;13:307.
85. Fick DM, Cooper JW, Wade WE, *et al*. Updating the Beers criteria for potentially inappropriate medication use in older adults: results of a US consensus panel of experts. *Arch Intern Med* 2003;163:2716–24.
86. Gallagher P, Ryan C, Byrne S, *et al*. STOPP (Screening Tool of Older Person's Prescriptions) and START (Screening Tool to Alert doctors to Right Treatment). Consensus validation. *Int J Clin Pharmacol Ther* 2008;46:72–83.
87. Schafer I, von Leitner EC, Schon G, *et al*. Multimorbidity patterns in the elderly: a new approach of disease clustering identifies complex interrelations between chronic conditions. *PLoS ONE* 2010;5:e15941.