BMJ Open Partially randomised patient preference trials as an alternative design to randomised controlled trials: systematic review and meta-analyses

Karin A Wasmann 1, Pieta Wijsman, Susan van Dieren, Willem Bemelman, Christianne Buskens

ABSTRACT

Objective Randomised controlled trials (RCT) are the gold standard to provide unbiased data. However, when patients have a treatment preference, randomisation may influence participation and outcomes (eg, external and internal validity). The aim of this study was to assess the influence of patients’ preference in RCTs by analysing partially randomised patient preference trials (RPPT); an RCT and preference cohort combined.

Design Systematic review and meta-analyses.

Data sources MEDLINE, Embase, PsycINFO and the Cochrane Library.

Eligibility criteria for selecting studies RPPTs published between January 2005 and October 2018 reporting on allocation of patients to randomised and preference cohorts were included.

Data extraction and synthesis Two independent reviewers extracted data. The main outcomes were the difference in external validity (participation and baseline characteristics) and internal validity (lost to follow-up, crossover and the primary outcome) between the randomised and the preference cohort within each RPPT, compared in a meta-regression using a Wald test. Risk of bias was not assessed, as no quality assessment for RPPTs has yet been developed.

Results In total, 117 of 3734 identified articles met screening criteria and 44 were eligible (24873 patients). The participation rate in RPPTs was >95% in 14 trials (range: 48%–100%) and the randomisation refusal rate was >50% in 26 trials (range: 19%–99%). Higher education, female, older age, race and prior experience with one treatment arm were characteristics of patients declining randomisation. The lost to follow-up and crossover rate were significantly higher in the randomised cohort compared with the preference cohort. Following the meta-analysis, the reported primary outcomes were comparable between both cohorts of the RPPTs, mean difference 0.093 (95% CI −0.178 to 0.364, p=0.502).

Conclusions Patients’ preference led to a substantial proportion of a specific patient group refusing randomisation, while it did not influence the primary outcome within an RPPT. Therefore, RPPTs could increase external validity without compromising the internal validity compared with RCTs.

PROSPERO registration number CRD42019094438.

Strengths and limitations of this study

- This systematic review and meta-analyses of partially randomised patient preference trials (RPPT) provides unique data on external and internal validity between randomised and patients’ preference cohorts.
- It was not possible to objectively establish the quality of included trials, as there is currently no valid critical appraisal tool to apply for an RPPT.
- Uniform counselling is of crucial importance in RPPTs, which has not been standardly reported in the included studies.

INTRODUCTION

Randomised controlled trials (RCT) are suggested to provide the most reliable evidence for treatment efficacy. However, participants are not passive recipients of interventions. Patients with a treatment preference may decline enrolment to avoid being randomised to their non-preferred treatment. Consequently, treatment preferences can decrease the generalisability of RCT results to the clinical population (ie, reduce external validity). Additionally, trials comparing experimental versus standard treatment are likely to include patients preferring experimental treatment, as trial participation is not needed for patients preferring standard treatment, further reducing external validity. Internal validity may be reduced, as randomisation to the (non-) preferred strategy could influence adherence to treatment protocol and study outcomes. Subjective study outcomes can directly be affected by treatment preference, whereas objective outcomes are most likely affected indirectly via adherence (so-called reluctant acquiescence phenomenon). Especially for an unblinded trial comparing treatments of significant different nature (eg, medical vs surgical) the RCT could be
an inappropriate design. Throughout the years, several approaches using various names have been proposed as alternative designs to diminish the influence of patients’ preference on validity: a partially randomised patient preference trial (RPPT), a comprehensive cohort trial, a patient preference trial, and more. In general, the aim of these designs is to treat patients with a preference for treatment strategies accordingly, whereas only those patients without a distinct preference will be randomised in the usual way. In the era of patients becoming more active participants in research, the use of RPPTs increases. The two previous systematic reviews addressing influence of preference on validity concluded that this influence was limited. However, one review only included studies addressing psychotherapy, and the other dates from 2005. So far, the value of the RPPT remains unclear, nor has it been addressed in the Oxford Levels of Evidence (Centre for Evidence-Based Medicine).

The aim of the study was to assess the influence of patients’ preference following randomisation in current daily clinical practice, by comparing randomised cohorts with preference cohorts within all RPPTs published since 2005. Two hypotheses were tested: (1) Patients’ preference will negatively influence participation in RCTs, decreasing external validity. Therefore, the external validity of an RPPT will be higher. (2) Patients’ preferences will influence adherence and outcomes in RCTs, decreasing internal validity. However, as only the remaining indifferent patients will be included in the RCT cohort of an RPPT, this RCT cohort can be considered as the true gold standard for internal validity.

METHODS
Design
A systematic review and meta-analyses of RPPTs was conducted. This study is reported in accordance with the Cochrane Handbook for Systematic Reviews of Interventions and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (online supplementary material 1). The study protocol is available in online supplementary material 2.

Data sources and searches
A search in PubMed, Embase, PsycINFO and the Cochrane Library for RPPTs published between 1 January 2005 and 5 October 2018 was executed without language restriction with the assistance of a librarian. The subject in the search strategy was RPPT and possible aliases of RPPT (see the PubMed Search Strategy section). Database searches were supplemented by hand searching reference lists of relevant articles. Additionally, authors were contacted to seek for data from unpublished studies identified. Non-English language articles were translated for possible inclusion.

PubMed search strategy
5 October 2018


And


“comprehensive cohort*[tiab] AND (“2004/09”[Date - Publication] : “3000”[Date - Publication])

Study selection
RPPTs describing results of both the randomised and preference cohorts, as long as in both cohorts patients met the same inclusion and exclusion criteria and were treated according to the same treatment protocol, were included. Trials in which a two-stage randomised design was conducted, allocation was based on doctors’ preference, without available separate data for the randomised and preference cohorts, with economic primary outcomes, or with non-clinical populations were excluded. Furthermore, it was decided not to include older RPPTs (before 2005), as it is important to consider the value of this design for current daily practice. A previous systematic review addressing the value of RPPTs was published in 2005, which can be used to interpret results from older studies.

Data extraction
The two first authors independently screened the citations and abstracts for eligible articles using a prepiloted standardised data form (Covidence; Veritas Health Innovation, Melbourne, VIC, Australia). Disagreements were discussed at steering group meetings.

The same two authors extracted data with the use of the same data form. Multiple publications reporting on the same trial were considered as one single trial for these analyses.
The level of sought data was summary estimates. Authors were contacted for further information when necessary. In case they were not forthcoming, the study was included in the review, but excluded from our reanalysis and/or meta-analyses.

Risk of bias assessment
Quality assessment of the trials was not performed, as no quality assessment for RPPTs has yet been developed and current criteria predominantly relate to concealment of randomisation (eg, Risk of Bias in Non-Randomized Studies-I and Cochrane Risk of Bias); consequently quality assessment and variability between trials were not applicable. Since the outcomes of each trial greatly differed, also the risk of bias assessment for systematic reviews (eg, Grading of Recommendations, Assessment, Development and Evaluations) was not applicable.

Outcomes
The primary outcomes were external and internal validity between randomised and preference cohorts within RPPTs. To analyse whether patients’ preference influenced external validity, data were extracted on participation rates in the randomised and preference cohorts. To assess if a specific patient group accepted randomisation, data were extracted on baseline characteristics of the randomised and preference cohorts of an RPPT separately. These characteristics were categorised into sociodemographic and clinical factors. Subsequently, these factors were compared between the randomised and preference cohorts of RPPTs.

To analyse whether patients’ preference influenced internal validity, data were extracted on lost to follow-up, crossovers and primary outcomes of the randomised and preference cohorts of an RPPT separately. Subsequently, these outcomes were compared between the randomised and preference cohorts within RPPTs. The primary outcomes of RPPTs were identified through explicit statements, study hypotheses, reported power analyses, and were checked on similarity with the study protocol. If this was not sufficient, the most likely primary outcome was chosen by consensus (KAW and SvD), or the study was excluded. To compare the primary outcomes between the randomised and preference cohorts within RPPTs, the outcome effects were compared between the randomised cohort and the preference cohort. It is emphasised that comparisons of outcome between randomised and preference cohorts are subject to bias, and if not done by the study itself, it was not possible to adjust for confounding factors. If in studies the adjusted and non-adjusted primary outcomes were available, the adjusted outcomes were used. Subsequently, separate analyses on adjusted and non-adjusted primary outcomes were performed.

Statistical analysis
The randomisation rate, participation rate and difference in baseline characteristics between the randomised and preference cohorts were explored and described, but not compared using statistics. To assess differences in baseline characteristics, mean and SDs were compared. If median IQRs were reported, it was converted to mean and SDs. When baseline characteristics were presented per experimental and control group, the sum of mean and SDs of these two groups was calculated for the randomised and preference cohorts using a weighted t-test. The lost to follow-up and cross-over rates were compared using a random effects model meta-analysis for proportions.

To realise the comparison of the primary outcomes of randomised and preference cohorts, a reanalysis was conducted. Because the trials involved a range of diseases, outcome measures and sample sizes, different treatment effect scales were converted into standardised effect sizes in the reanalysis. Treatment effects were calculated directly for continuous outcome variables as standardised mean differences (difference in means divided by the pooled SD). For binary outcomes, log ORs were calculated and converted into standardised effect size differences. In case none of the patients in the preference cohort chose the control treatment, the treatment effect of the experimental treatment was compared with the control treatment of the randomised cohort. Only trials for which a ‘net’ effect (primary outcome minus baseline value of the primary outcome) could be calculated were included in the meta-analyses. In case the ‘net’ effect was missing, but baseline values and primary outcomes were available, the SD was estimated. Heterogeneity was not assessed as trial outcomes were different for each study included. Meta-analysis of randomised versus preference cohort was performed using a random effects model with an inverse variance weighting. A final meta-regression was performed using a Wald test to compare the standardised treatment effects.

A p<0.05 was considered a significant difference. R’s programming environment was used (V.3.5.1, R Foundation for Statistical Computing, Vienna, Austria).

Patient and public involvement
There was no direct involvement of patients or the public in the development of the research question, selection of the outcome measures, design and implementation of the study, or interpretation of the results.

RESULTS
In total, 117 out of 3734 records identified were full text screened. Fifty-eight partially RPPTs from 2005 onwards were found, of which 44 (including 24 873 patients) were eligible for at least basic data extraction (table 1), and 20 could be included in the meta-analyses (PRISMA flow chart, figure 1). Exclusion reasons for the meta-analyses were: no availability of both treatment outcomes in the randomised and preference cohorts separately in 14 trials, no availability of SDs, which could also not be converted from other
<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>R (n)</th>
<th>P (n)</th>
<th>Field</th>
<th>Intervention and comparison groups</th>
<th>Primary outcome(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashok et al</td>
<td>Women presenting for termination of pregnancy</td>
<td>400</td>
<td>86</td>
<td>Gynaecology</td>
<td>Medical versus surgical termination†</td>
<td>Acceptability at 2 weeks</td>
</tr>
<tr>
<td>Barnard et al</td>
<td>Premenopausal women with symptomatic uterine fibroids</td>
<td>59</td>
<td>34</td>
<td>Gynaecology</td>
<td>UAE versus MrGFUS†</td>
<td>Perioperative outcomes at 3 months</td>
</tr>
<tr>
<td>Bergk et al</td>
<td>Patients with DSM-IV disorder</td>
<td>27</td>
<td>81</td>
<td>Psychiatry</td>
<td>Mechanical restraint versus seclusion</td>
<td>CES at 4 weeks</td>
</tr>
<tr>
<td>Boers et al</td>
<td>Pregnant women with disproportional intrauterine growth</td>
<td>650</td>
<td>452</td>
<td>Gynaecology</td>
<td>Induction versus expectative monitoring*</td>
<td>(SAE neonate at discharge)</td>
</tr>
<tr>
<td>Brinkhaus et al</td>
<td>Patients with allergic rhinitis</td>
<td>981</td>
<td>4256</td>
<td>Social medicine</td>
<td>Acupuncture versus control*</td>
<td>AQLQ at 3 months</td>
</tr>
<tr>
<td>Brinkhaus et al</td>
<td>Patients with allergic rhinitis</td>
<td>981</td>
<td>4256</td>
<td>Social medicine</td>
<td>Acupuncture versus control*</td>
<td>RQLQ at 30 days</td>
</tr>
<tr>
<td>Buhagiar et al</td>
<td>Patients after total knee arthroplasty</td>
<td>195</td>
<td>87</td>
<td>Orthopaedics</td>
<td>Home-based versus inpatient recovery</td>
<td>Walking distance at 36 weeks</td>
</tr>
<tr>
<td>Chekherov et al</td>
<td>Elderly with ovarian cancer receiving chemotherapy</td>
<td>3</td>
<td>116</td>
<td>Gynaecology</td>
<td>Oral versus intravenous treosulfan</td>
<td>DFS at 2 years</td>
</tr>
<tr>
<td>Creutzig et al</td>
<td>Paediatric patients with relapsed AML</td>
<td>101</td>
<td>54</td>
<td>Haematology</td>
<td>L-DNR/FLAG versus FLAG</td>
<td>OS at 4 years</td>
</tr>
<tr>
<td>Crowther et al</td>
<td>Pregnant women with one prior caesarean</td>
<td>22</td>
<td>2332</td>
<td>Gynaecology</td>
<td>Caesarean versus vaginal birth†</td>
<td>Death and SAE at 30 days</td>
</tr>
<tr>
<td>Dalal et al</td>
<td>Participants in cardiac rehabilitation after acute MI</td>
<td>104</td>
<td>128</td>
<td>Cardiology</td>
<td>Home-based versus hospital recovery</td>
<td>HADS at 9 months</td>
</tr>
<tr>
<td>Ejlersen et al</td>
<td>Premenopausal patients with breast cancer</td>
<td>525</td>
<td>1628</td>
<td>Oncology</td>
<td>Chemotherapy versus ovarian ablation†</td>
<td>DFS at 10 years</td>
</tr>
<tr>
<td>Erkan et al</td>
<td>Patients with positive aPL but no vascular and/or pregnancy events.</td>
<td>98</td>
<td>74</td>
<td>Internal medicine</td>
<td>Aspirin versus placebo or no aspirin*</td>
<td>Acute thrombosis per 100 patient-years</td>
</tr>
<tr>
<td>Fong et al</td>
<td>Patients with adolescent idiopathic scoliosis</td>
<td>19</td>
<td>50</td>
<td>Orthopaedics</td>
<td>Brace versus observational*</td>
<td>Recruitment feasibility</td>
</tr>
<tr>
<td>Gall et al</td>
<td>Patients undergoing colon cancer surgery</td>
<td>203</td>
<td>135</td>
<td>Surgery</td>
<td>GP versus surgeon follow-up</td>
<td>PCS score at 24 months</td>
</tr>
<tr>
<td>Glazener et al</td>
<td>Patients with vaginal wall prolapse</td>
<td>1348</td>
<td>1126</td>
<td>Gynaecology</td>
<td>Mesh versus no mesh†</td>
<td>POPSS at 12 months</td>
</tr>
<tr>
<td>Grant et al</td>
<td>Patients with gastro-oesophageal reflux disease</td>
<td>357</td>
<td>453</td>
<td>Upper GI</td>
<td>Surgery versus medication†</td>
<td>Reflux QoL at 1 year</td>
</tr>
<tr>
<td>Hatcher et al</td>
<td>Patients presenting with self-harm</td>
<td>552</td>
<td>542</td>
<td>Psychiatry</td>
<td>PST plus standard care versus standard care*</td>
<td>Repeated self-harm at 1 year</td>
</tr>
<tr>
<td>Howard et al</td>
<td>Women requiring voluntary psychiatric admission</td>
<td>42</td>
<td>61</td>
<td>Psychiatry</td>
<td>Crisis houses versus psychiatric wards</td>
<td>Functioning (GAF) at 12 weeks</td>
</tr>
<tr>
<td>Hubacher et al</td>
<td>Women aged 18–29 years who were seeking a short-acting method</td>
<td>382</td>
<td>524</td>
<td>Gynaecology</td>
<td>Long-acting versus short-acting contraceptive*</td>
<td>Continuation rate at 1 year</td>
</tr>
<tr>
<td>Jones et al</td>
<td>Patients with palliative cancer</td>
<td>41</td>
<td>36</td>
<td>Oncology</td>
<td>Advance versus usual care*</td>
<td>VAS(S) at 8 weeks</td>
</tr>
<tr>
<td>Karlsen et al</td>
<td>Patients with proximal ureter stones</td>
<td>50</td>
<td>21</td>
<td>Urology</td>
<td>Shock wave versus ureteroscopy†</td>
<td>Stone-free rate at 3 months</td>
</tr>
<tr>
<td>Kearney et al</td>
<td>Patients with an acute Achilles tendon rupture</td>
<td>20</td>
<td>29</td>
<td>Orthopaedics</td>
<td>Surgery versus conservative†</td>
<td>Disability rating index at 9 months</td>
</tr>
<tr>
<td>Kröz et al</td>
<td>Patients with breast cancer-related fatigue</td>
<td>65</td>
<td>61</td>
<td>Oncology</td>
<td>Multimodal combined programme versus aerobic training*</td>
<td>PSQI at 10 weeks</td>
</tr>
<tr>
<td>Lock et al</td>
<td>Children with recurrent sore throats</td>
<td>268</td>
<td>461</td>
<td>Children surgery</td>
<td>Surgery versus medication†</td>
<td>Number of episodes of sore throat at 2 years</td>
</tr>
<tr>
<td>Majumdar et al</td>
<td>Patients with lower urinary tract symptoms (LUTS)</td>
<td>99</td>
<td>210</td>
<td>Urology</td>
<td>Urodynamics versus conservative†</td>
<td>King's QoL at 6 months</td>
</tr>
<tr>
<td>Mittal et al</td>
<td>Patients with type B ankle fracture</td>
<td>160</td>
<td>276</td>
<td>Orthopaedics</td>
<td>Surgery versus no surgery†</td>
<td>FAQO and PCI at 12 months</td>
</tr>
<tr>
<td>Prescott et al</td>
<td>Women after breast-conserving surgery</td>
<td>255</td>
<td>100</td>
<td>Oncology</td>
<td>Non-radiotherapy versus radiotherapy*</td>
<td>QoL after 5 years</td>
</tr>
<tr>
<td>Purepong et al</td>
<td>Office workers suffering from low back pain (LBP)</td>
<td>64</td>
<td>37</td>
<td>Physical therapy</td>
<td>Backrest versus no intervention*</td>
<td>VAS at 3 months</td>
</tr>
<tr>
<td>Raue et al</td>
<td>Patients operated for diverticulitis</td>
<td>149</td>
<td>294</td>
<td>Surgery</td>
<td>Laparoscopic versus open approach</td>
<td>QoL at 30 days</td>
</tr>
<tr>
<td>Robson et al</td>
<td>Termination of pregnancy less than 14 weeks' gestation</td>
<td>349</td>
<td>1528</td>
<td>Gynaecology</td>
<td>Medicine versus surgery TOP†</td>
<td>Acceptability TOP at 2 weeks</td>
</tr>
<tr>
<td>Schweikert et al</td>
<td>Patients for cardiac rehabilitation</td>
<td>4</td>
<td>163</td>
<td>Cardiology</td>
<td>Outpatient versus inpatient recovery</td>
<td>EQ-5D at 12 months</td>
</tr>
<tr>
<td>Shi et al</td>
<td>Patients with vascular dementia</td>
<td>48</td>
<td>20</td>
<td>Complementary medicine</td>
<td>Acupuncture versus training*</td>
<td>SDSVD at 6 weeks</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>R (n)</th>
<th>P (n)</th>
<th>Field</th>
<th>Intervention and comparison groups</th>
<th>Primary outcome(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinclair et al.‡</td>
<td>Patients with severe lung disease</td>
<td>67</td>
<td>82</td>
<td>Pulmonology</td>
<td>Advance care planning versus standard</td>
<td>ACP uptake at 6 months</td>
</tr>
<tr>
<td>Schwegler et al.‡</td>
<td>Adolescent with idiopathic scoliosis (AIS)</td>
<td>132</td>
<td>187</td>
<td>Orthopaedics</td>
<td>Brace versus observation</td>
<td>QoL at 2 years</td>
</tr>
<tr>
<td>Underwood et al.§‡</td>
<td>Older patients with chronic knee pain</td>
<td>282</td>
<td>303</td>
<td>Orthopaedics</td>
<td>Topical versus oral ibuprofen</td>
<td>WOMAC at 12 months</td>
</tr>
<tr>
<td>Underwood et al.§‡</td>
<td>Adolescent with idiopathic scoliosis (AIS)</td>
<td>177</td>
<td>103</td>
<td>Gynaecology</td>
<td>Embolisation versus hysterectomy†</td>
<td>HRQoL at 12 months</td>
</tr>
<tr>
<td>Underwood et al.§‡</td>
<td>Older patients with chronic knee pain</td>
<td>29</td>
<td>10</td>
<td>Orthopaedics</td>
<td>Surgery versus botulinum therapy†</td>
<td>SHUEE at 24 weeks</td>
</tr>
<tr>
<td>Weiner et al.‡</td>
<td>Patients with spondylolisthesis</td>
<td>304</td>
<td>303</td>
<td>Orthopaedics</td>
<td>Surgical versus non-surgical†</td>
<td>Physical functioning (SF-36 Phys) at 2 years</td>
</tr>
<tr>
<td>Van Heest et al.‡</td>
<td>Patients with spinal stenosis</td>
<td>289</td>
<td>365</td>
<td>Orthopaedics</td>
<td>Surgical versus non-surgical†</td>
<td>Physical functioning (SF-36 Phys) at 2 years</td>
</tr>
<tr>
<td>Wittbrodt et al.‡‡</td>
<td>Addicted people</td>
<td>293</td>
<td>321</td>
<td>Social medicine</td>
<td>Community residential versus day hospital*</td>
<td>Abstinence at 12 months</td>
</tr>
<tr>
<td>Witt et al.‡‡</td>
<td>Patients with chronic low back pain</td>
<td>2841</td>
<td>8537</td>
<td>Rheumatology</td>
<td>Acupuncture versus control*</td>
<td>HFAQ at 3 months</td>
</tr>
<tr>
<td>Witt et al.‡‡</td>
<td>Patients with osteoarthritis</td>
<td>712</td>
<td>2921</td>
<td>Rheumatology</td>
<td>Acupuncture versus control*</td>
<td>Osteoarthritis index (WOMAC) at 3 months</td>
</tr>
<tr>
<td>Woodward and Kelly‡‡</td>
<td>Pregnant women</td>
<td>60</td>
<td>20</td>
<td>Gynaecology</td>
<td>Water versus land birth</td>
<td>Baby condition at 6 weeks</td>
</tr>
</tbody>
</table>

*These 32 trials compared interventions versus conservative treatment.
†These 16 trials compared surgical interventions versus conservative treatment.
‡These 20 trials could be used to calculate standardised effect sizes of the randomised and preference cohorts separately, and were included in our reanalysis on the effect of preference on outcome.

ACP, advance care planning; aPL, antiphospholipid; AQLQ, Asthma Quality of Life Questionnaire; CES, Coercion Experience Scale; DFS, disease-free survival; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; EQ-5D, EuroQol-5 Dimension; FAQ, Foot and Ankle Outcomes Questionnaire; FLAQ, Fudarabin; GAF, Global Assessment of Functioning; GP, general practitioner; HADS, Hospital Anxiety Depression Scale; HFAQ, Hannover Functional Ability Questionnaire; HRQoL, health-related quality of life; L-DNR, Lipoosomal daunorubicin; MII, myocardial infarction; MRgFUS, MRI-guided focused ultrasound surgery; OS, overall survival; P, preference; PG, physical component score; PCS, peripheral cancer score; PSQI, Pittsburgh Sleep Quality Index; PST, problem-solving therapy; R, randomised; ROLQ, Rhinois Quality of Life Questionnaire; SAE, serious adverse event; SDFSD, scale of differentiation of syndromes of vascular dementia; SF-36, Short Form-36; SHUEE, Shriners Hospital Upper Extremity Evaluation; TIP, termination of pregnancy; UAE, uterine artery embolisation; VAS, visual analogue scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.
available data in five trials, and the number of events or the power of one or both cohort(s) was too low to perform separate randomised and preference analyses in five trials. The trials covered a wide range of clinical areas and interventions. The main areas were gynaecology (n=11), orthopaedics (n=10) and psychiatry (n=5). Of the 44 included trials, 32 compared an intervention versus conservative treatment, including 16 surgical interventions (table 1). In all trials but one, if patients refused randomisation they received their preference treatment (figure 2). In the other study, a Zelen randomisation was performed, randomising all eligible patients and afterwards asking for their consent to participate in the randomised arm or if they preferred the other intervention. Parental preference was relevant in five trials involving children, as permission of parents was required and the preference between patients and parents could not be distinguished.

External validity
The following results concern the influence of patients’ preference on external validity. Information on the number of eligible patients who agreed to participate (in either the randomised or preference cohort) was available in 39 out of the 44 RPPTs. The participation rate of eligible patients in the RPPTs ranged from 48% to 100%, in which 16 RPPTs reported a participation rate higher than 80%, and 14 RPPTs with a participation rate higher than 95%. Of these included participants in the 44 RPPTs, 18%-99% declined randomisation (hence these patients were included in the preference cohort). The randomisation refusal rate was more than 50% in 26 RPPTs.

To assess if a specific patient group accepted randomisation, 35 of the 44 RPPTs reported at least one comparison between randomised and preference cohorts on baseline sociodemographic factors. At least one significant difference between randomised and preference cohorts was found in 20 of the 35 trials. Overall, 38 significant differences were found in 161 sociodemographic comparisons (24%). The proportion of significant findings was not dependent on sample size (smaller trials n<300; 19/85, 22% and larger trials n≥300; 19/76, 25%). Patients with a preference compared with those accepting randomisation were more likely to be older, female, with higher education, employed, Caucasian, not obese, non-smokers, unmarried and experienced with one treatment arm (online supplementary material 3).

Thirty-four of the 44 RPPTs reported at least one comparison between randomised and preference cohorts on clinical baseline characteristics. At least one significant difference was found in 20 of the 34 trials. Overall, 36 significant differences were found in 220 clinical comparisons (16%). The proportion of significant findings was not dependent on sample size (smaller trials n<300; 12/78, 15% and larger trials n≥300; 24/142, 17%). Patients with a preference had more severe clinical problems in seven trials and less severe clinical problems in 10 trials, while in the remaining three trials no consistent pattern could be found (online supplementary material 3).
Internal validity

The following results concern the influence of patients’ preference on internal validity. Information on lost to follow-up in both the randomised and preference cohorts was available in 33 of the 44 RPPTs. For the randomised cohorts, the proportion of individuals lost to follow-up was <10% in 14 trials, 10% to <20% in 9 trials and ≥20% in 10 trials. For the preference cohorts the corresponding numbers of trials were 17, 9 and 7. The mean percentage of participants lost to follow-up was significantly higher in the randomised cohorts (16.1%, SD 16.8%) compared with the preference cohorts (13.3%, SD 14.7%), relative risk (RR 1.3) (95% CI 1.0 to 1.6, p=0.03).

Information on crossovers in both the randomised and preference cohorts was available in 20 of 44 RPPTs. For the randomised cohorts, the proportion of individuals who crossed over to the other study treatment was <10% in 11 trials, 10% to <20% in 5 trials and ≥20% in 4 trials. For the preference cohorts the corresponding numbers of trials were 14, 5 and 1. The mean percentage of crossovers was significantly higher in the randomised cohorts (14.5%, SD 16.9%) compared with the preference cohorts (6.3%, SD 11.5%), RR 2.6 (95% CI 1.7 to 3.9, p<0.001).

To assess the influence of patients’ preference on primary outcomes, for 20 of the 44 RPPTs it was possible to perform reanalyses using standardised effect sizes (figure 1).

Figure 3 shows the magnitude of the experimental treatment effect over the control treatment effect of the randomised and preference cohorts separately using standardised effect sizes. The trials are listed by sample size. A positive experimental treatment effect was seen in 13...
trials. The influence of patients’ preference on primary outcomes according to different standardised treatment effects between randomised and preference cohorts was small; in 13 of the 20 trials (65%) this was 0.2 or less (scale −2 to 2), in 5 trials (25%) between 0.21 and 0.5, and in 2 trials (10%) higher than 0.5. Of the 20 RPPTs, the overall mean difference in primary outcome between randomised and preference cohorts was not significantly different, 0.093 (95% CI −0.178 to 0.364, p=0.502) (figure 2). Only two trials showed a significant different treatment effect between the randomised and preference cohorts. In both trials the experimental treatment effect was favourable over the control treatment effect in both the randomised and preference cohorts, but the favourable effect of the experimental treatment was significantly greater in the preference cohort. Both RPPTs compared acupuncture versus conservative treatment. In one trial the improvement of the osteoarthritis index in patients with osteoarthritis of the knee or hip was assessed, the other trial assessed the functional ability score in patients with chronic low back pain.

In 7 of these 20 trials, an adjusted primary outcome for baseline confounders was available. In these trials, the mean difference in primary outcome between randomised and preference cohorts was even smaller, −0.026 (95% CI −0.263 to 0.211, p=0.832). In 18 trials (also) a non-adjusted primary outcome was available. Using these outcomes, the mean difference in primary outcomes was 0.228 (95% CI −0.117 to 0.572, p=0.196) (figures 4 and 5).

**DISCUSSION**

These study results challenge the current consensus about the hierarchy of study designs. Our results indicate that patients’ preference led to a substantial proportion of patients refusing randomisation (refusal of randomisation was more than 50% in 26 trials), while it did not affect the primary outcome of an RPPT.

Regarding our first hypothesis, it can be concluded that patients’ preference does negatively influence participation to RCTs, as demonstrated by the low participation to the randomised cohort in RPPTs. The participation in the RPPTs was remarkably high (ranging from 48% to 100%), improving external validity when compared with the classic RCT (ranging from <0.001% to 40%).

Cautiously, it could be argued that a typical patient group characterised by, for example, higher education, Caucasian race and non-obese individuals are more likely to refuse randomisation. In contrast, differences in clinical characteristics showed no consistent pattern in the randomised or preference cohorts. Therefore, not including a patient’s preference cohort in a trial could result in a potential loss of inclusions of a specific patient group, further decreasing external validity.

Regarding our second hypothesis, it can be concluded that patients’ preference does not significantly affect the primary outcome of an RPPT, as the primary outcomes of patients in the randomised and preference cohorts were similar. Since the aim of an RPPT is to treat patients according to their preference, it can be assumed that the randomised cohort of an RPPT includes patients indifferent to the type of treatment. Subsequently, it is unlikely that outcomes of randomised patients will be biased by treatment preference. Hence, they could be seen as the gold standard. Lost to follow-up and crossovers were significantly higher in the randomised cohort compared with the preference cohort. As a result, the data of the preference cohort could be interpreted more easily than

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>N</th>
<th>P. outcome</th>
<th>Favours control</th>
<th>Favours experimental</th>
<th>Std effect size (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones et al</td>
<td>Randomised</td>
<td>37</td>
<td>Discussion VAS</td>
<td>-</td>
<td>-</td>
<td>0.05 (-0.59; 0.70)</td>
</tr>
<tr>
<td>Howard et al</td>
<td>Preference</td>
<td>31</td>
<td>-</td>
<td>0.11 (-0.62; 0.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buhagiar et al</td>
<td>Randomised</td>
<td>165</td>
<td>Functioning</td>
<td>-</td>
<td>-</td>
<td>0.11 (-0.86; 0.65)</td>
</tr>
<tr>
<td>Underwood et al</td>
<td>Randomised</td>
<td>246</td>
<td>Walking distance</td>
<td>-</td>
<td>-</td>
<td>0.01 (-0.30; 0.31)</td>
</tr>
<tr>
<td>Weinstein et al</td>
<td>Preference</td>
<td>254</td>
<td>Osteoarthritis Index</td>
<td>-</td>
<td>-</td>
<td>-0.12 (-0.42; 0.19)</td>
</tr>
<tr>
<td>Weinstein et al</td>
<td>Randomised</td>
<td>252</td>
<td>Functioning</td>
<td>-</td>
<td>-</td>
<td>0.02 (-0.30; 0.25)</td>
</tr>
<tr>
<td>Grant et al</td>
<td>Randomised</td>
<td>299</td>
<td>Reflux QoL</td>
<td>0.11 (-0.12; 0.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>Preference</td>
<td>321</td>
<td>-</td>
<td>0.37 (-0.14; 0.60)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4** Forest plot of the preference effect on the primary outcome between the randomised and preference cohorts of trials in which the primary outcome is adjusted for confounders. The overall treatment effect (standardised effect size) within the randomised cohorts was compared with the overall treatment effect within the preference cohorts. QoL, quality of life; RCT, randomised controlled trial; PP, Patients’ preference cohort; VAS, visual analogue scale.
Figure 5  Forest plot of the preference effect on the primary outcome between the randomised and preference cohorts of trials in which the primary outcome is not adjusted for confounders. The overall treatment effect (standardised effect size) within the randomised cohorts was compared with the overall treatment effect within the preference cohorts. ACP, advance care planning; AQLQ, Asthma Quality of Life Questionnaire; FAQO, Foot and Ankle Outcomes Questionnaire; HADS, Hospital Anxiety Depression Scale; HFAQ, Hannover Functional Ability Questionnaire; PP, Patients’ preference cohort; QoL, quality of life; RCT, randomised controlled trial; SDSVD, scale of differentiation of syndromes of vascular dementia; VAS, visual analogue scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

downloaded from http://bmjopen.bmj.com/ BMJ Open: first published as 10.1136/bmjopen-2019-031151 on 16 October 2019. Downloaded from 

the randomised data. Perhaps, consciously choosing a treatment ensures a certain dedication and tolerance for the treatment.

Our results are strengthened by the previous systematic review of King et al, including RPPTs from 1966 to 2004. Based on their results, they also postulated that treatment preference influences the willingness to accept randomisation, and that the evidence of its significant effect on internal validity is low.4 A possible limitation of their study is that they did not measure patients’ preference as specifically as in our analyses, since they also included a minority of two-stage randomised trials, as physician preference.

An RCT is once designed to reliably compare medication to placebo.74 In the hierarchy of research designs, the results of RCTs are considered to be evidence of the highest grade. Lessons learnt from the history of RCT, and early studies from 1970s and 1980s suggested that observational studies suffer too much from confounders and frequently result in overestimation of treatment effects compared with RCTs.75 76 Consequently, many experts advocated that results of observational studies should not be used for defining evidence-based medical care: ‘If the study wasn’t randomized, we suggest that you stop reading it and go on to the next article.’77 However, two updates of this work including studies between 1985 and 1995 found little evidence that estimates of treatment effects in observational studies are consistently larger than those obtained in RCTs.78 79 It is suggested that observational studies have methodologically improved over time with the use of a control group, carefully defining inclusion and exclusion criteria, and by better understanding confounders. The fundamental criticism of the RPPT could be that within the preference cohort the unrecognised confounding factors may distort the results. Yet, our results showed that preference cohorts provide valid information comparable with the randomised results.

Today, the classic levels of evidence are subject of debate, as the disadvantages of RCTs have become more insightful in modern practice. In general, patients participating in RCTs are highly selected. Less than 10% of patients participate in trials, partly due to exclusion of patients with a specific treatment preference.80 This limits the extrapolation of RCT results to patients seen in routine practice. Another consequence is that the majority of trials take several years to be completed. This causes a burden on health research costs, and results in a questionable ethical dilemma. Developments are fast and
the relevance of trials may therefore change over time. Consequently, if an RCT is optimally designed but takes too long, the results will be outdated.

This especially applies when designing a trial in which it can be foreseen that patients’ preference will be a prominent factor, for example, in trials comparing treatments of significant different nature (medical vs surgical). Anticipation on the expected patients’ preference by eliminating this factor is at the expense of the validity of a lot of RCTs. Especially when patient-centred outcomes are used, one should consider whether the most important patient group has been excluded. Trials must be internally valid, but lack of consideration of external validity causes the widespread undue of treatments—that showed superiority in RCTs—in routine practice. Moreover, in these situations an RPPT could be the superior design over an RCT.

RPPTs provide unique data on external and internal validity as the patients in the preference cohort are followed according to the same conditions as the patients in the randomised cohorts. A limitation of our review is that interventions and settings between RPPTs were very diverse. On the other hand, because of this diversity, it could also be stated that randomised and preference data often produce similar results in all kinds of settings. Concerning the assessment of external validity, it should be noted that in only a minority of trials the differences in sociodemographic and clinical parameters between the cohorts of an RPPT were evident. Furthermore, in some cases none of the patients in the preference cohort choose the control treatment. In these cases, the treatment effect of the experimental treatment was compared with the control treatment of the randomised cohort. These are not optimal comparisons, but considered to be more appropriate than excluding these data. Moreover, as the idea of RPPTs is a relatively new concept, various terms were used in the inclusion period of this systematic review. In the publication of Walter et al in 2017, different concepts were compared and they clearly defined the terms fully randomised patient preference trial and partially randomised patient preference trial. To achieve a ‘fully randomised patient preference trial’, the preference of all participants should be identified. Therefore, uniform counselling is of crucial importance in RPPTs. The majority of included studies claim to be RPPTs. However, in most of currently included studies, the details of how patients were counselled have not been addressed. As we cannot guarantee that a study identified the preference of all eligible patients, we decided to use the term partially randomised patient preference trials. Another result of the novelty of such a design is that it was not possible to objectively establish the quality of included trials, as there is currently no valid critical appraisal tool to apply for an RPPT. Consequently, our results may have been influenced by the inclusions of flawed trials. In conclusion, RPPTs seem to be a reliable alternative for RCTs, especially in trials comparing treatments of vastly different nature (eg, medical vs surgical) or using patient-centred outcomes. In case patients’ preference can be assumed, RPPT enables faster inclusion of a more representative population improving external validity without compromising internal validity.

Contributors KW and CB designed the study. KW and PW performed the search. KW and SvO did the statistical analyses. KW wrote the first draft with input of CB and WB.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iD Karin A Wasmann http://orcid.org/0000-0003-2295-238X

REFERENCES
14. 16.1.3.2 imputing standard deviations for changes from baseline. Available: http://handbook-5.1.cochrane.org/chapter_16/16_1_3_2_imputing_standard_deviations_for_changes_from_baseline.html [Accessed 10 Jul 2018].


73 Rothwell PM. External validity of randomised controlled trials: “To whom do the results of this trial apply?” The Lancet 2005;365:82–93.


