Objective
To propose an improvement to the PRECIS-2 scoring method, by providing a scoring grid for retrospective scoring of surgical and pharmaceutical pragmatic clinical trials.

Background
The PRECIS-2 tool was developed to support the design of pragmatic clinical trials, and is widely used for trial evaluation in health care research. However, recent studies have highlighted the limitations of the tool.

The PRECIS-2 tool was designed to facilitate the design of pragmatic trials, but may not be suitable for evaluating existing trials.

Methods
Development of the scoring grid

The scoring grid (criteria for each of the 9 domains, for each of the 2 types of trial) was defined using guidelines from the PRECIS-2 publication and the example studies provided by the PRECIS-2 authors.

Structured literature review

A structured literature review was conducted in PubMed (search terms: “pragmatic trial” AND “trial design”) with the publication dates limited to the past 5 years.

Studies with a single-arm clinical trial were selected, as they are used in surgical or pharmaceutical settings. Non-patient, observational studies were excluded. A total of 341 studies and 10 pharmaceutical interventions were included.

Included studies were assessed according to the PRECIS-2 scoring grid by a primary scorer, followed by an independent assessment by a validation scorer.

Analysis

A “Study Score” was calculated for each study, based on the sum of scores from the 9 domains.

Average “Study Scores” were calculated for the surgical and pharmaceutical studies.

In order to analyse the consistency in scoring between the primary scorers and the validation scorers, the absolute differences were calculated and averaged across all the domains and studies (average absolute difference, AAD).

Results

- Of the 341 search results, 6 surgical and 10 pharmaceutical interventions were included.
- Results from the primary scoring suggest the surgical studies and the pharmaceutical studies were similar in terms of degree of pragmatism, with overlapping “Study Score” distributions (mean SD 4.15 (0.49) and 4.22 (0.49), respectively).
- Domain 6 (Flexibility of experimental intervention) had the most commonly un-scored domain due to lack of information.
- The most pragmatic domains were domain 3 (Setting, average score 4.83) and domain 9 (Analysis, average score 4.70) for the surgical and pharmaceutical studies, respectively.
- The results from the analysis of the primary scores were confirmed by the validation scores, which also suggested the surgical and pharmaceutical studies were similar.
- The AAD was much larger for the pharmaceutical studies (0.753) than for the surgical studies (0.1875) and in both study types the primary scores were significantly different from the validation scores (paired t test, p<0.001) (Figure 1C and D).

Discussion

- Our analysis did not identify a large difference in the degree of pragmatism between surgical studies and pharmaceutical studies.
- The “Study Score” results should nonetheless be interpreted with care, due to inconsistencies in scoring as well as the following limitations:
  - Selection bias may have been introduced by our literature review eligibility criteria which excluded combination interventions.
  - The choice of a combination intervention may be part of a pragmatic approach to clinical trial design.
  - Publication bias is likely to have been introduced, since only sufficiently credible pragmatic trials would have been accepted by peer reviewers.
  - The equal weighting of all the domains, and the assignment of criteria along the 1–5 scale, might not necessarily reflect the relative importance of each of the domains for pragmatic clinical trial designs.

- Even though the differences between the primary scores and the secondary scores were significant (p<0.001), these differences might not be meaningful given the low AAD, especially for surgical studies (AAD=0.1875).
- The AAD of the pharmaceutical studies was higher than the AAD of the surgical studies, suggesting that the scoring grid was less well-adapted for scoring of pharmaceutical trials. Other possible reasons for the observed variation include:
  - Poor reporting: This issue was particularly relevant for domain 4 (organisation), as it was rarely explicitly explained whether additional training would be required for health care professionals (HCPs). For a “generalist” score, it is not always clear if the HCPs would already have sufficient skills from standard training to deliver the interventions.
  - Variation in benefit of the doubt: Some scorers might be more likely to give the benefit of the doubt when information is incomplete, leading to variations in the scores.

Conclusions

- Our initial research suggests that a scoring grid for PRECIS-2 could assist objective scoring of clinical trials on the pragmatic-exploratory study design continuum. However, the scoring grid should be further refined in order to ensure more consistent scoring across a range of disease areas and interventions.

References