Efficacy of interventions for the treatment of irritable bowel syndrome, functional abdominal pain—not otherwise specified, and abdominal migraine in children: a systematic review and network meta-analysis

Vasiliki Sinopoulou, Jip Groen*, Morris Gordon, Ed Mougey, James P Franciosi, Tim G J de Meij, Merit M Tabbers, Marc A Benninga*

Summary

Background Many treatments for abdominal pain-related disorders of gut–brain interaction (AP-DGBI) in children have been studied. We aimed to assess the efficacy and safety of all known treatment options for paediatric AP-DGBI.

Methods For this systematic review and network meta-analysis, we searched Embase, MEDLINE, and CENTRAL databases from inception to Jan 16, 2025, for published randomised controlled trials. We included trials of any treatment for AP-DGBIs (irritable bowel syndrome, functional abdominal pain—not otherwise specified, and abdominal migraine, excluding functional dyspepsia) in children aged 4–18 years. We excluded randomised controlled trials that solely included children with functional dyspepsia, but we included studies in which children with functional dyspepsia were included alongside children with the other AP-DGBI diagnoses and outcome data could not be separated. Data extraction and quality appraisal were performed in duplicate. The primary outcome for this network meta-analysis was author-defined treatment success. Network meta-analysis methodology was used within a frequentist framework using multivariate meta-analysis and outcomes were assessed using the Grading of Recommendations, Assessment, Development and Evaluation methodology. Clinical relevance of effect sizes was interpreted according to consensus definitions.

Findings Of 19 337 records identified through the database search, 155 records representing 91 original randomised controlled trials were included in the network meta-analysis: these 91 trials comprised 7226 participants (4119 females and 2673 males). 12 studies assessed dietary treatments (n=730), 25 assessed pharmacological treatments (n=2140), 23 assessed probiotic treatments (n=1762), and 35 assessed psychosocial treatments (n=2952). Two treatments were probably more effective for treatment success than control treatments (moderate certainty): hypnotherapy (risk ratio [RR] 4·99 [95% CI 2·15 to 11·57]; large effect size) and cognitive behavioural therapy (CBT; RR 1·99 [95% CI 1·33 to 2·98]; moderate effect size). All other treatments evaluated for treatment success were either not effective or the data were of very low certainty and thus no conclusions could be made.

Interpretation Hypnotherapy and CBT show moderate certainty for treatment efficacy with clinically relevant effect sizes. No conclusions can be made about the other therapies and treatment success due to very low evidence certainty. Future randomised controlled trials should focus on improving the evidence certainty for those other therapies with regard to core AP-DGBI outcomes.

Funding None.

Lancet Child Adolesc Health 2025; 9: 315–24

*Joint first authors

School of Medicine, University of Central Lancashire, Preston, UK (V Sinopoulou RD, Prof M Gordon PhD); **Department of Pediatric Gastroenterology** (J Groen MD, T G J de Meij PhD, M M Tabbers PhD, Prof M A Benninga PhD), **Department of Gastroenterology Endocrinology Metabolism** (J Groen, M M Tabbers), **and Reproduction and Development Research Institute** (J Groen, M M Tabbers), **Amsterdam University Medical Center, Amsterdam, Netherlands; Clinical Pharmacogenomics and Translational Research, Nemours Children's Health System, Jacksonville, FL, USA** (E Mougey PhD); **Department of Pediatrics, Division Gastroenterology, Hepatology, and Nutrition, School of Medicine, University of Louisville, Louisville, KY, USA** (J P Franciosi MD)

Correspondence to:
Prof Morris Gordon, School of Medicine, University of Central Lancashire, Preston PR1 2HE, UK
mgordon@uclan.ac.uk

Introduction

Abdominal pain-related disorders of gut-brain interaction (AP-DGBI) have a global prevalence of 11·7% and cause chronic, debilitating pain in more than 300 million children annually worldwide.[1] Children with AP-DGBIs have lower health-related quality of life (similar to inflammatory bowel disease), more missed school days on average than their peers, and more frequent hospital admissions than healthy children, and up to a third will continue to have symptoms into adulthood.[2,3] The quality of evidence from randomised clinical trials and systematic reviews of treatment options for AP-DBGI has not been well characterised. Comprehensive Grading of Recommendations, Assessment, Development and Evaluation (GRADE) assessment for AP-DBGI treatments is needed to support paediatric health-care professionals in treatment decisions.

The Rome IV committee presents diagnostic criteria for four separate entities comprising AP-DGBI: (1) irritable bowel syndrome (IBS; prevalence of 5·8%); (2) functional abdominal pain—not otherwise specified (FAP-NOS; prevalence of 1·2%); (3) abdominal migraine (prevalence of 1·7%); and (4) functional dyspepsia

Research in context

Evidence before this study

We searched PubMed from database inception to Jan 16, 2025, using the search string ''(functional abdominal OR abdominal pain OR gut-brain) AND (child* OR pediatr* OR youth) AND treatment'', limiting the results to network meta-analyses. No language restrictions were applied. Selection criteria were network meta-analyses of randomised controlled trials examining any (set of) treatment options for abdominal pain disorders of gut–brain interaction (AP-DGBIs) in children (aged <18 years), in which outcomes related to any improvement in symptom severity were measured. Of the five network meta-analyses identified, only one examined the effects of treatment options for children with AP-DGBIs, but investigated dietary options only. This review did not use the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) to assess certainty of the effects, but rather used the Confidence in Network Meta-Analysis (CINeMA) approach. Certainty of the effect was assessed per comparison but not per outcome and assessment of imprecision in CINeMA is based on a single threshold. No previously published network meta-analyses have compared treatment categories beyond dietary options or combining multiple treatment categories (eg, pharmacological and dietary), which reflects an important evidence gap. In addition to treatment options with clear evidence of effect, the use of poorly studied options and best practice options is variable and considerable regional differences in prescription patterns exist.

Added value of this study

To our knowledge, this network meta-analysis presents the most comprehensive attempt to both synthesise and judge certainty of the evidence on all available treatment options for AP-DGBIs in children. A novel addition, following the 2013 GRADE guidance, was the prospective identification of clinical decision thresholds for interpretation of effect sizes, which has an advantage over interpreting effect size through statistical instruments. We investigated whether treatment efficacy is the result of an isolated intervention or is likely to include overlooked subcomponents, such as concomitant or standard care, but none of these factors impacted the efficacy of interventions in the network. The findings establish certainty of the efficacy of cognitive behavioural therapy and hypnotherapy, while there is no evidence for efficacy of other commonly used treatment options. The present network meta-analysis provides the most robust and up-to-date evidence-synthesis for the entirety of treatment options in AP-DGBIs for policy makers, guideline developers, affected children, and their families and carers.

Implications of all the available evidence

In terms of clinical guidance, this network meta-analysis supports the use of psychotherapies in AP-DGBIs, and highlights several other options that might have future benefit, the evidence for which should be strengthened by additional high-quality trials. This is reflected in the forthcoming international treatment guideline for these disorders in children due to be published by a collective of the European and North American Societies for Pediatric Gastroenterology, Hepatology and Nutrition. The meta-analysis highlights for authors of future trials on AP-DGBIs the importance of clearly describing all factors in treatment modules separate to the treatment itself. These should include consideration of concomitant care alongside the intervention of interest, educational offerings, dietary advice, and prognostic expectations. Prognostic expectations are highly confounded by the considerable range of control and placebo treatment efficacy in the present evidence, and the scarcity of objective tools of symptom assessment other than self-reporting.

(prevalence of 2·1%).[1,4] Similar management strategies have been adopted, specifically for IBS and FAP-NOS, which are often studied jointly in trials.[5–8]

Known treatment options for AP-DGBI can be divided into pharmacological, dietary, psychotherapies, probiotics, and percutaneous nerve field stimulation. Systematic reviews of paediatric trials have identified a variety of active interventions against non-active control, but there is a paucity of head-to-head comparisons between active interventions.[5–8] When accounting for the large placebo effect in children with AP-DGBI,[9] superiority of active interventions over control interventions is often uncertain and therefore direct comparison of active treatments might result in greater certainty of effects.

Network meta-analysis allows for direct and indirect comparison of interventions. To date, no network meta-analyses have been done to compare all types of AP-DGBI treatments. Considering the variety of treatment options, uncertainty about their efficacy and safety—as reflected by a paucity of treatment guidelines—and regional differences in prescribing patterns are common. A network meta-analysis could provide a thorough synthesis of the various treatments to support evidence-based management and evidence-informed choices among therapeutic options.[10] Network meta-analyses are often used primarily as a way to rank therapies, but are regularly done without clearly defining the outcomes of interest,[11,12] without GRADE certainty of evidence in the network,[13] and without using treatment effect thresholds to support assessment of statistical imprecision and clinical relevance.[14] Therefore, we aimed to perform a systematic review and network meta-analysis of randomised controlled trials to assess the efficacy and safety of all treatments in children with AP-DGBIs (IBS, FAP-NOS, and abdominal migraine) in accordance with GRADE[15] and Cochrane methodology.[16]

Methods

Search strategy and selection criteria

A prospective protocol for this systematic review with network meta-analysis is available on an open repository.[17] The requirement for ethical approval was waived since no original or individual data were obtained. The study adhered to Cochrane guidance for systematic reviews,[16] GRADE guidance for network meta-analyses,[13,15] and PRISMA reporting guidelines (appendix pp 179–81).[18]

The scope of the review included three of the four AP-DGBI conditions: IBS, FAP-NOS, and abdominal migraine. The full diagnostic criteria for each condition are included in the protocol.[17] We excluded the fourth AP-DGBI entity defined in the Rome IV diagnostic categories[4]—functional dyspepsia—in accordance with previous research in the field[10,19] based on it being a separate disease category with notably different treatment approaches and outcome measures. A systematic search was designed by a Cochrane information specialist (YY) who searched Embase, MEDLINE, and CENTRAL databases from inception to Jan 16, 2025, using the search terms as described in the appendix (pp 146–51). No language or other restrictions were applied.

We included all randomised controlled trials that compared any treatments with active treatment, placebo, standard care, or no treatment (waitlist with no other active therapy), in children aged 4–18 years with AP-DGBI, specifically IBS, FAP-NOS, or abdominal migraine, as defined by the authors using the Rome criteria. Randomised controlled trials that solely included children with functional dyspepsia were excluded, but studies in which children with functional dyspepsia were included alongside the other AP-DGBI diagnoses and outcome data could not be separated, were included. Cross-over randomised controlled trials were included, but only data collected before crossover were used when available. We contacted the original authors of studies if needed for clarification or additional data to aid our screening, risk of bias assessments, and analyses.

Titles and abstracts were screened in duplicate by three authors (JG, MG, and VS) and three acknowledged contributors (SL, AA, and DANA). Full-text screening was performed in duplicate by JG, MG, and VS. Any disagreements were resolved by consensus agreement between authors or by a third author (MB, MT, and VS).

Data analysis

The primary outcome of interest was treatment success, as defined by the authors (dichotomous outcome). Secondary outcomes of interest were abdominal pain frequency or change in frequency of pain using any validated scale (continuous outcome); abdominal pain intensity or change in pain intensity using any validated scale (continuous outcome); and serious adverse events as defined by the authors (dichotomous outcome). These outcomes were selected because they were identified as critical outcomes (as per GRADE assessment) during the development of the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) and North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN) guidelines for AP-DGBIs.[10]

We used clinical thresholds for the interpretation of the magnitude of effect sizes. These thresholds were prospectively determined through an online Delphi process as part of the ESPGHAN and NASPGHAN guideline development process (table; appendix p 95).[10] The Delphi process required clinical experts to identify clinically relevant thresholds (eg, 10% more treatment success for one treatment versus another) and assign categories (ie, trivial, small, moderate, or large clinical effect) to these thresholds to assess whether a statistically significant difference between a treatment and its comparator was also clinically significant. Computing effect size is commonly performed using statistical instruments, such as Cohen's *d*, however, such metrics partly ignore whether the effect size is clinically relevant. The Delphi process was completed by a group of multi-professional experts in the field before commencing the review.

See Online for appendix

Data were extracted using a standardised data extraction form. Extracted data included demographic and baseline characteristics, intervention details, and outcome data. Risk of bias was assessed using the Cochrane risk of bias 1 tool.[20] JG, VS, and EM performed data extraction and risk of bias was assessed independently in duplicate, with any disagreements resolved by an additional author (MG). Outcome data for the end of intervention timepoints were extracted.

The authors discussed and agreed by consensus which treatments were sufficiently homogenous to be grouped

	Large magnitude effect size	Moderate magnitude effect size	Small magnitude effect size	Trivial magnitude effect size
Treatment success, as defined by authors (dichotomous), %	>40%	25 to 40%	10–24%	<10%
Pain frequency reduction in episodes per week (continuous)	>12	8 to 12	4 to <8	<4
Pain intensity reduction (measured using VAS [score 0–10]; continuous)	>2·6	1·5 to 2·6	0·7 to <1·5	<0·7
Quality of life increase (measured using PedsQL [score 0–100]; continuous)	>42	25 to 42	11 to <25	<11
Stool consistency improvement (measured using BSS [score 1–7]; continuous)	>2·5	1·6 to 2·5	0·8 to <1·6	<0·8
Withdrawals due to adverse events or serious adverse events (dichotomous), %	>4%	3 to 4%	1 to <3%	<1%
Total adverse events (dichotomous), %	>16%	8 to 16%	4 to <8%	<4%

Percentages indicate risk difference from inactive control treatments. Thresholds for interpretation of effect sizes were determined through a Delphi consensus process for guideline development by the European Society for Paediatric Gastroenterology, Hepatology and Nutrition and the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition. AP-DGBI=abdominal pain-related disorders of gut–brain interaction. BSS=Bristol Stool Scale. PedsQL=Pediatric Quality of Life Inventory. VAS=visual analogue scale.

Table: **Clinically relevant interpretation of magnitude of effect for study outcomes related to AP-DBGI**

together as single treatment for the network meta-analysis. For the main network analysis, the control treatments placebo, waitlist, and all forms of standard care, or no intervention, were grouped together as an umbrella control treatment. Control treatment was used as the index therapy to which all others were compared.

Dichotomous outcomes were expressed as risk ratios (RR) with corresponding 95% CIs, and risk difference between treatments and control treatments was calculated and expressed as percentages. Continuous outcomes were expressed as mean differences with 95% CIs. For continuous outcomes assessed on more than one scale, we estimated internal reference SDs and change from baseline mean (SD) values using a correlation coefficient of 0·5, and standardised outcome results as change from baseline on the most commonly used outcome scale.[21] RR (95% CI), risk difference, and mean difference (95% CI) are presented for all treatments in the summary of findings tables and for all treatments with GRADE certainty of low or better in graphical plots. Analyses were done by intention-to-treat. We used a random-effects model to pool data.

Network meta-analysis methodology was employed within a frequentist framework using multivariate meta-analysis.[16] We assessed the assumption of transitivity (the assumption that population characteristics and other important factors are similar across the included studies) by comparing the distribution of potential effect modifiers across pairwise comparisons. Heterogeneity was assessed statistically using the I^2 statistic for each pairwise comparison, and with the loop-specific approach for the direct and indirect estimates. Surface under the cumulative ranking curve (SUCRA), combined with GRADE, was used to rank treatments. Funnel plots were used to assess publication bias for pairwise analyses with at least ten studies. Component network meta-analysis was employed to investigate the effect of separate intervention and control components.[22] Component network meta-analysis can help to identify whether there is an additive effect of intervention components that are applied in different combinations throughout the randomised controlled trials evidence base, for example, when comparing the following intervention types: hypnotherapy plus standard care, placebo plus standard care, and standalone standard care. In a regular network meta-analysis, these would be considered three completely different interventions, whereas component network meta-analysis recognises the shared component of standard care in the three groups. Statistical analyses were performed using R statistical software (version 4.5.0) and netmeta package.

We did several preplanned subgroup analyses: (1) a control analysis where the control treatments of placebo, waitlist and standard care, or no intervention were considered as separate treatments (referred to as the split control analysis hereafter); (2) subdiagnosis analysis (IBS *vs* FAP); and (3) analysis by age group (4–12 years *vs* 13–18 years).

We also did several pre-planned sensitivity analyses: (1) a random versus fixed-effects statistical model; (2) a component network meta-analysis; (3) per diagnostic criteria (eg, Rome criteria iterations *vs* Apley's criteria); (4) per outcome definition (only applicable to treatment success); and (5) removal of studies deemed to have a high risk of bias. Two additional post-hoc sensitivity analyses were performed for the outcome of treatment success: removal of studies including participants with functional dyspepsia and removal of studies where success was defined based on quality of life or social functioning improvement (due to heterogeneity with the other definitions used for treatment success).

The GRADE framework was used to assess the certainty of the evidence.[15] Since all studies were randomised controlled trials, the outcomes were initially assessed as high certainty. We then assessed direct and indirect evidence certainty based on GRADE risk of bias, inconsistency, indirectness, and publication bias. The network evidence certainty was also assessed based on imprecision and incoherence between direct and indirect evidence. Two authors (MG and VS) independently rated the certainty ratings and disagreements were resolved by discussion and consensus with the wider team. The evidence was rated as high, moderate, low, or very low. Results for analysis were presented using a GRADE Of Results Diagram Of Network meta-analysis plot,[23] which represents the magnitude and certainty of results ranked by magnitude of effect within a given certainty class. When outcomes were of very low certainty, meaning conclusions should not be drawn (regardless of the magnitude or absolute effects observed), they were not included in these plots.

GRADE was used in combination with SUCRA to rank treatments. In the summary of findings tables treatments were ranked from high to low SUCRA probability and their corresponding GRADE certainty and estimates were presented. Treatments were presented from high to low GRADE certainty and ranked by SUCRA probability within their respective GRADE assessment rating (high, moderate, or low).

Role of the funding source

There was no funding source for this study.

Results

We included a total of 91 randomised controlled trials (figure 1; appendix pp 5–19, 22–33, 152–78) including 7226 participants. 249 children with functional dyspepsia were included in mixed populations within the studies and could not be removed from outcome data (3·4%). An additional five randomised controlled trials (n=273) included children with functional dyspepsia, but did not specify how many (appendix pp 20–21).

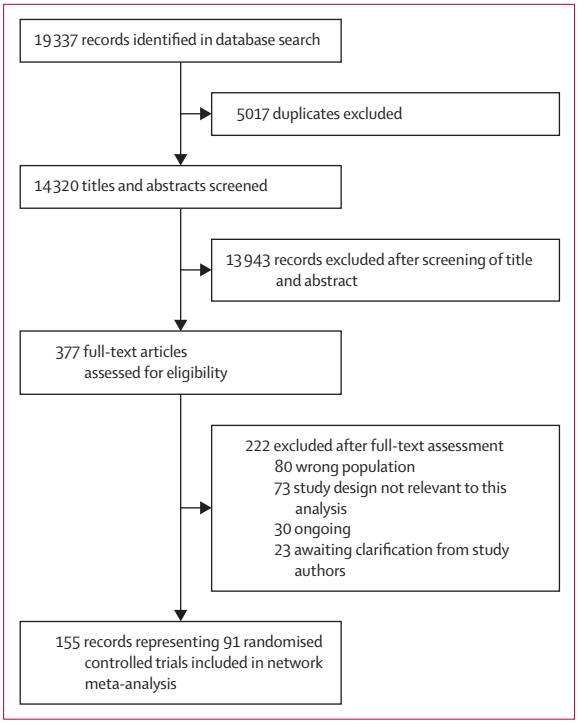

Figure 1: **PRISMA flow diagram**

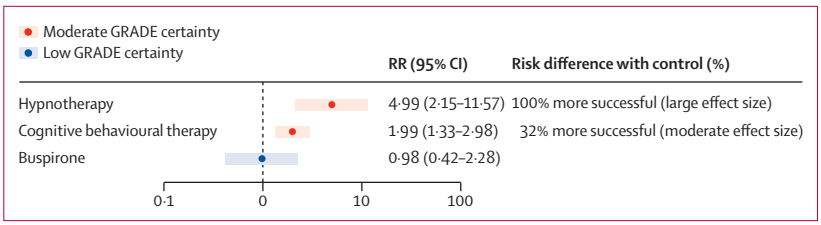

Figure 2: **Graphical plot of treatment success network meta-analysis results compared with control**
Comparison of treatment success (treatment *vs* inactive control treatments) assessed with a random-effects model. Treatments with very low GRADE certainty are not included. RR=risk ratio. GRADE= Grading of Recommendations, Assessment, Development and Evaluation.

Overall, 12 trials of dietary interventions (n=730; 416 [57%] females and 241 [33%] males), 25 trials of pharmacological interventions (n=2140; 1155 [54%] females and 984 [46%] males), 23 trials of probiotic interventions (n=1762; 969 [55%] females and 792 [45%] males), and 35 trials of psychosocial interventions (n=2952; 2155 [73%] females and 797 [27%] males) were included in our network meta-analysis.

85 (93%) of the 91 included studies reported the sex of study participants (4119 [61%] of 6792 participants were female and 2673 [39%] of 6792 participants were male).

Only 17 (19%) of 91 included studies reported race or ethnicity. In those studies, 1207 (70%) of 1724 participants were White, 242 (14%) participants were Asian, 59 (3%) participants were Black, 57 (3%) participants were mixed race, 25 (1%) were Hispanic, and 143 (8%) were of other ethnicity. No studies reported subgroup outcome data based on sex, age groups, or race and ethnicity.

Age, sex, and race or ethnicity were similar across the included studies, supporting that the assumption of transitivity holds (ie, the assumption that population characteristics and other important factors are similar across the included studies).

The risk of bias assessment summary for all the included studies is presented in the appendix (pp 34–94).

58 studies reported treatment success as a dichotomous outcome (appendix pp 96–97). 31 studies defined treatment success based on predefined reductions in pain frequency, duration, intensity or severity, or a combination of these factors. 11 studies used complete resolution of pain as the definition of treatment success. 11 studies used adequate relief, consistent with the previously published core outcome set for childhood FAPD disorders,[24] and three studies used quality of life or function as the focus for their success score. 48 studies (n=3846) were connected in the main network meta-analysis for treatment success, comparing a total of 18 treatments (figure 2; appendix pp 116–35). The overall control success rate for the combined control treatments was 322 successes per 1000 children (range 0–785 successes per 1000 children). No treatments were rated as high certainty. Two treatments, hypnotherapy (RR 4·99 [95% CI 2·15–11·57]; large effect size) and cognitive behavioural therapy (CBT; 1·99 [1·33–2·98]; moderate effect size) were of moderate certainty, suggesting they are probably more effective for treatment success than the control treatments (figure 2). Buspirone was of low certainty suggesting no difference compared with the control (0·98 [0·42 to 2·28]) for treatment success (figure 2). All other treatments were rated at very low certainty and therefore no conclusions could be made (appendix pp 104–05). A subgroup analysis in which the control treatments were separated into placebo, waitlist and standard care, and no intervention, with placebo plus standard care as the comparison treatment (419 successes per 1000 children), showed differences compared with the main analysis, with increased imprecision throughout the network (appendix pp 117–36). Subgroup analysis per subdiagnosis and age group were not possible due to the large heterogeneity of the included subdiagnoses and age groups in the included studies, which did not allow for connected networks, and paucity of subgroup outcome data. A sensitivity analysis with a fixed-effect model identified no major differences compared with the main analysis (appendix p 141). The sensitivity analyses by diagnostic criteria, treatment success definitions (appendix pp 22–24), and risk of bias were not possible because no homogeneous connected networks could be established. The results of the component network meta-analysis sensitivity analysis were substantially different from the results of the main network meta-analysis (appendix pp 117–36), and the subgroup network meta-analysis for separated control treatments (appendix pp 117–36), but as this separated the network into large numbers of small and imprecise comparisons, no

conclusions could be drawn. We also conducted two post-hoc sensitivity analyses. In the first analysis, we excluded all studies that included children with functional dyspepsia and in the second analysis, we excluded all studies that based their treatment success definitions on improvements in quality of life and social functioning (appendix p 142–43). The results were similar to the main analysis, whereby the same treatments were identified as effective. A publication bias funnel plot could only be performed for the nine randomised controlled trials of probiotic interventions. Visual inspection did not identify major concerns for publication bias (appendix p 140).

50 studies provided outcome data for pain intensity using a variety of pain scales (appendix pp 98–100). 45 studies (n=3187 participants) were connected in the main network meta-analysis for pain intensity, comparing a total of 19 treatments (appendix pp 106–10, 118–37). All treatments were rated at very low certainty and therefore no conclusions could be drawn. A subgroup analysis in which the control treatments were separated into placebo, waitlist, standard care, and no intervention, with placebo plus standard care as the comparison treatment, showed no major differences compared with the main analysis (appendix pp 119–36). Subgroup analyses per subdiagnosis and age group were not possible. A sensitivity analysis with a common (fixed effect) model identified no major differences compared with the main random-effects analysis (appendix p 144). However, results of the component network meta-analysis sensitivity analysis were substantially different from the main network meta-analysis (appendix pp 119–38), and the subgroup network meta-analysis for separate control treatments, but since this separated the network into large numbers of small and imprecise comparisons no conclusions could be drawn. The sensitivity analyses by diagnostic criteria and risk of bias were not possible since no homogeneous connected networks could be established. A publication bias funnel plot could only be done for the 12 trials of probiotic interventions. Visual inspection did not identify any major concerns for publication bias (appendix p 140).

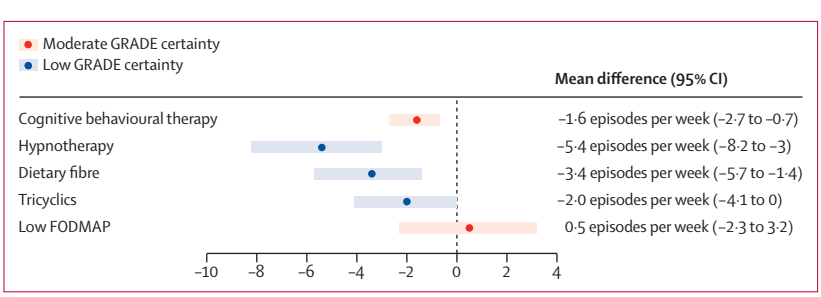

Figure 3: **Graphical plot of pain frequency network meta-analysis results compared with control as differences in episodes per week**

Comparison of pain frequency (treatment *vs* inactive control treatments) assessed with a random-effects model. Treatments with very low GRADE certainty were not included. GRADE= Grading of Recommendations, Assessment, Development and Evaluation. FODMAP=fermentable oligosaccharides, disaccharides, monosaccharides, and polyols.

25 studies (n=1870) provided outcome data for pain frequency (appendix pp 100–01) and were connected in the main network meta-analysis for pain frequency, comparing a total of 12 treatments (figure 3; appendix pp 111–14, 120–139). As per GRADE assessment, no treatments were of high certainty. Two treatments were deemed to be of moderate certainty; CBT (mean difference 1·6 fewer episodes per week [95% CI 0·7–2·7]; trivial effect) was probably more effective than the combined control treatments at reducing pain frequency from baseline in episodes per week; and the low fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAP) diet is probably no different to the combined control interventions at reducing pain frequency from baseline (mean difference 0·5 more episodes per week [95% CI 2·3 fewer episodes to 3·2 more episodes]). Three treatments were of low certainty; hypnotherapy (mean difference 5·4 fewer episodes per week [95% CI 3·0–8·2]; small effect) and dietary fibre (mean difference 3·4 fewer episodes per week [1·4–5·7]; trivial effect) were possibly more effective than the combined control treatments at reducing pain frequency from baseline in episodes per week, while tricyclic antidepressants (mean difference 2·0 fewer episodes per week [0–4·1]) were probably no different to the combined control interventions (figure 3). All other treatments were rated as very low certainty and thus no conclusions could be drawn (appendix pp 113–14). A subgroup analysis in which the control treatments were separated into placebo, waitlist and standard care, and no intervention, with placebo plus standard care as the comparison treatment, resulted into two subnetworks, which showed no major differences compared with the main analysis (appendix pp 121–39). Subgroup analyses per subdiagnosis and age group were not possible. In the sensitivity analysis with a common fixed-effects model, no major differences were identified compared with the main analysis (appendix p 145). The results of the component network meta-analysis sensitivity analysis was identical to the subgroup analyses of separated controls (appendix pp 121–39). The sensitivity analyses by diagnostic criteria and risk of bias were not possible. A publication bias funnel plot could only be performed for eight of the 12 trials of probiotic interventions. Visual inspection did not indicate publication bias (appendix p 140). Network plots for all efficacy outcome networks are shown in figure 4, which shows the number and types of direct comparisons within the networks.

No network meta-analysis was possible for serious adverse events, since safety reporting in the included studies was sparse and for most studies that did report safety outcomes, no serious adverse events occurred (appendix pp 101–03). Only three serious adverse events were reported in all studies combined. Two adverse events occurred in a study by Di Lorenzo and colleagues, on linaclotide versus placebo, with one occurring in both study groups (appendix p 153). The nature of these

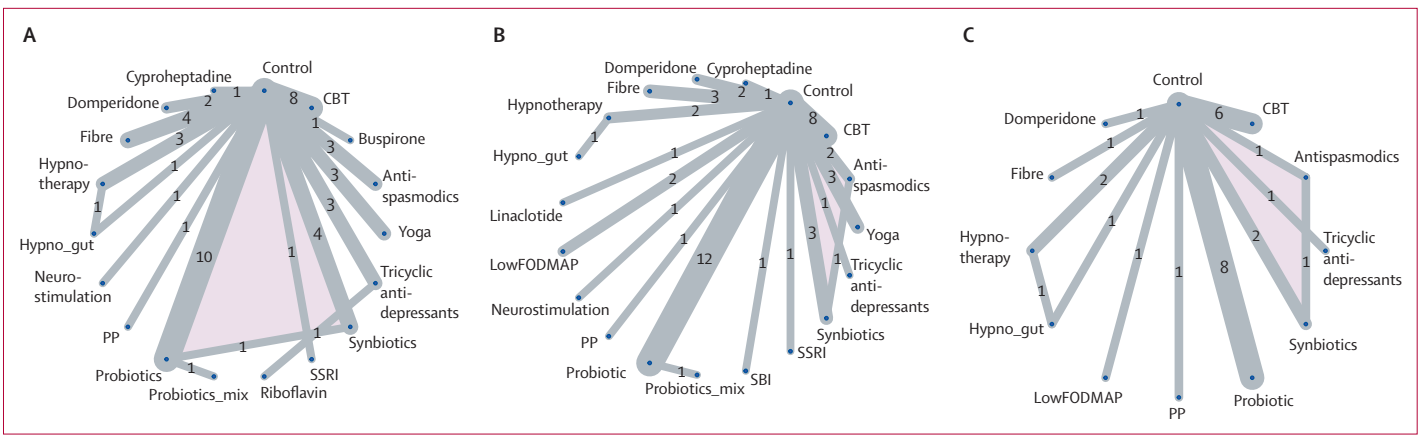

Figure 4: **Network meta-analysis plots for treatment success (A), pain intensity (B), and pain frequency (C)**
Treatments with direct comparisons are linked with a line; the thickness of the line corresponds to the weight of the random-effects model comparing the two treatments. Numbers on connecting lines correspond to the number of trials comparing the two treatments. Purple shaded areas show the largest directly connected networks that were not compared with the index therapy. CBT=cognitive behavioural therapy. FODMAP=fermentable oligosaccharides, disaccharides, monosaccharides, and polyols. Hypno_gut=gut-directed hypnotherapy. Probiotics_mix=probiotic preparations with strain mixtures. PP=Palmitoylethanolamide and polydatin. SBI= serum-derived bovine immunoglobulin. SSRI=Selective serotonin reuptake inhibitors.

events was not specified, however the authors reported they were unrelated to the study medication. The other serious adverse event occurred in a study by Vázquez-Frias and colleagues that compared *Bacillus claussi* with placebo (appendix p 161). One participant in the probiotic group had febrile multisystem inflammatory syndrome that lasted 2 days and was of moderate severity. The authors assessed this was unrelated to the study treatment.

Discussion

To our knowledge, this is the first comprehensive network meta-analysis of randomised controlled trials of interventions for paediatric AP-DGBI. Network meta-analysis as an approach is uniquely placed within paediatric AP-DGBI to produce high utility information by combining the data for control groups from trials to address the substantial variation in control efficacy effect sizes[9] and increase the validity of the findings from data with good certainty. The results from our network meta-analysis show that hypnotherapy and CBT probably have higher rates of treatment success than control interventions (moderate certainty), and are associated with clinically relevant, large and moderate effect sizes, respectively. Additionally, hypnotherapy might be more effective than control interventions for reducing pain frequency. CBT is probably more effective and dietary fibres are possibly more effective for reducing pain frequency than control treatments, albeit the possible effect sizes seem negligible. No meaningful conclusions could be reached with regard to pain intensity, since all interventions were of very low GRADE certainty, or intervention safety due to a paucity of reported serious adverse events across all studies.

Methodological limitations leading to low or very low certainty GRADE ratings were pervasive at both the direct meta-analysis and indirect network levels. Regardless of the magnitude of effect sizes, where certainty was very low no conclusions can be drawn. This prevented any conclusions being made about some classes or types of therapies (eg, for tricyclic antidepressants or antispasmodics) or in some cases different forms of therapies that are already within the network with higher certainty outcome data could not be considered, such as gut directed hypnotherapy (where suggestion is focused on the gut) as opposed to hypnotherapy. Imprecision was a pervasive issue that was noted in many of the very low certainty findings and is likely a function of the small sample sizes which is unfortunately common in the field.[25,26] Risk of bias was common and was largely related to deficiencies in reporting and the deficiencies in reporting were more pronounced in this review than in other areas of gastroenterology.[27] Despite detailed and repeated attempts to contact authors to clarify reporting gaps, responses were rarely received, which is not consistent with studies on author responses to reviewer data requests.[28] Statistical heterogeneity was common, despite our best efforts to account for differences in clinical and methodological factors within the analysis. Success of control group treatment ranged from 0% to 79%. Our preplanned subgroup analysis attempted to explain these differences by separating the various controls (no intervention, placebo, placebo with standard care, wait list) but this did not reduce the significant heterogeneity in control group effect sizes or significantly change the outcomes of the wider network meta-analysis findings. Standard treatment, educational offerings, dietary advice, previous treatments, and prognostic expectations were rarely mentioned in any form within the trials and we believe could all be key to understanding this complexity across trials. It is also possible that heterogeneity in the specific outcomes, such as treatment success, could have contributed to study heterogeneity,

although these were broadly based on combinations of similar key outcomes and this rendered it not possible to explore the impact further. A sensitivity analysis in which we removed studies that based their treatment success definitions on quality of life and improvements in social functioning, were similar to the main analysis. Some of the therapies were not suited to masking, such as CBT or hypnotherapy resulting in being downgraded for risk of bias. In line with accepted practice in the field,[29] in our analysis we were less stringent in downgrading GRADE assessments when the intervention could not be masked. However, this does mean that there is a potential confounding effect that could contribute to the unexplained heterogeneity.[6]

Despite these issues, efficacy of clinically relevant magnitude was observed for several interventions, which potentially form a core set of intervention options for the treatment of AP-DGBI. Indeed, hypnotherapy and CBT have been recommended as treatment options for AP-DGBI in revisions of the forthcoming ESPGHAN and NASPGHAN clinical guidelines and our study provides synthesised evidence to support these recommendations and the use of these interventions in clinical practice.

The findings of this network meta-analysis are broadly consistent with previous evidence syntheses in the field.[5–8] The authors of the current study have performed several systematic reviews and meta-analyses as precursors to this review and to inform the international guideline process.[5–8,10] The efficacy findings and certainty of evidence for CBT and hypnotherapy found in this study, is mirrored in one of these analyses that focused on psychosocial interventions.[6] Conversely, the paucity of evidence for efficacy of most dietary interventions has been previously recognised,[5] with probiotics being a possible exception. The results of a 2023 Cochrane systematic review and meta-analysis[7] are generally in agreement with our findings for probiotics, however a detailed subgroup analysis in the Cochrane study did find some efficacy for specific strains or mixtures. This observation is tempered by the low certainty of the evidence presented in the Cochrane study and by the lack of consideration of the clinical relevance of the magnitude of effect, and thus the clinical relevance must be interpreted with caution. Subgroup analysis at the network level of probiotic preparations in the current study was not possible due to imprecision concerns.[7]

The strengths of the current study include the innovative use of predefined effect size thresholds to judge imprecision and clinical meaningfulness, and the use of GRADE to assess the certainty of the results. Conventionally, standardised statistic strategies are used to determine effect size, such as Cohen's *d*. The predefined threshold approach, a feature that has been implemented in regular meta-analysis since guidance was published in 2022,[30] aims to represent effect size thresholds that have relevance from a clinical standpoint as opposed to a statistical standpoint, which further enhances the validity and utility of the findings in practice. The use of GRADE ensured a detailed accounting of the certainty of treatment effect. We did not place emphasis on an overall ranking of findings based on the network meta-analysis. Many network meta-analyses focus on the ranking as a key finding however, conventional network meta-analysis ranking is statistical (based on SUCRA or p value) and not advised as it does not account for clinical and contextual factors.[31] For example, a top ranking therapy based on SUCRA might not be indicated if the magnitude of the result is not of clinical relevance. Similarly, ranking does not consider GRADE and, as observed in this network meta-analysis, a top ranking is of no relevance if the results are of very low certainty. For example, the SUCRA results for the main analyses of our three efficacy outcomes (appendix pp 135, 137, 139) show treatments such as a specific probiotic mix, cyproheptadine, neurostimulation, tricyclics, domperidone, low FODMAP diet, dietary fibre, and others which are of very low certainty, ranking higher than hypnotherapy and CBT, which have shown the highest GRADE certainty of evidence among all network treatments. Thus, SUCRA rankings can easily lead to incorrect interpretations if not combined with GRADE, since they solely rely on statistical calculations, and do not account for imprecision, inconsistency, risk of bias, publication bias, or indirectness. Therefore, we believe the approach of combining GRADE and SUCRA should become standard. It is also not suggested that network meta-analyses are used in isolation for decision making and this analysis should not be considered superior to, but rather an adjunct to the direct pairwise analysis in the systematic reviews. Where the results are consistent between different evidence synthesis approaches, decision makers can be more certain in the data. This does not just support use in practice, but can also remove the need for future research in areas of certainty and focus these resources on evaluating less certain interventions.

This study has a number of limitations. The review does not include functional dyspepsia as a diagnostic category and the results of this review are not applicable to young people with functional dyspepsia. Within some studies included in this review, there were a small number of children with functional dyspepsia that could not be removed from mixed datasets. When we did account for this in a sensitivity analysis removing studies that included children with functional dyspepsia, we found no major differences compared with the main analysis results for treatment success. Attempts were made to examine the design of study groups in the evidence base, with a particular focus on the allowance of standard care alongside the proposed study treatment and a distinction between placebo, standard care, and waitlist. It was found that comparison groups are frequently poorly defined in study reports

and comparison group definitions can differ widely between study teams. Despite our attempts to expose such differences in group design to allow for a more accurate network meta-analysis, subgroup analyses based on factors such as the extent of concomitant (standard) care and elements of care that are not reported in control intervention elements, complicates interpretation of the network meta-analysis results due to increased imprecision. We were unable to involve people with lived experience of these conditions at any stage of our research, which is another limitation of this study.

An important conclusion of the current study is that descriptions of all interventions in trials, not just those being studied, should be standardised and described in sufficient detail to facilitate identification of factors that influence study outcomes. Such factors might include but are not limited to, concomitant therapies while on study treatment, care provider contact, educational strategies, and context of care focusing on the experience of care providers. Additionally, reporting of methods should directly address elements of bias so that they can be clearly assessed in accordance with international guidance,[32] since bias has a large influence on GRADE outcomes. Another important conclusion is the complete lack of subgroup outcome data in all included studies, which does not allow for any subgroup analyses by age, sex, or race and ethnicity. Consequently, questions on the role of these characteristics cannot be answered. Study authors are urged to publish relevant subgroup outcome data as this not only supports the interpretation of their own findings, but is valuable for future secondary analysis and systematic reviews which might support broader impact in the field.

Finally, these findings could inform future studies. The authors propose that further research of hypnotherapy and CBT is not a priority at this time given the certainty of evidence supporting their use. Instead, it is suggested that research focusses on key therapies in widespread use but with limited evidence to support their efficacy, which has been clearly highlighted by the very low certainty findings on multiple GRADE analyses. Examples include probiotic and dietary (eg, fibre) supplements and a number of pharmacological interventions and alternative therapies such as acupuncture or nerve field stimulation therapies. Investment in higher quality studies can ensure that certainty of findings increases, rather than just the volume of studies. Proper sample size estimation for adequate statistical power, and the young person or caretaker perspective on intervention and outcome prioritisation also need to be addressed in future trials.[24,33]

In summary, both hypnotherapy and CBT show moderate certainty evidence for treatment efficacy, suggesting they are probably effective for treating IBS, FAP-NOS, and abdominal migraine with clinically significant effect sizes. However, the majority of other interventions had very low-certainty evidence across outcomes and therefore no conclusions could be made. For those interventions, well-designed studies of adequate power are needed to determine efficacy.

Contributors

VS conceptualized the study, participated in data extraction, data analysis, writing of the manuscript and had full access to and verified the existence of the raw data. JG participated in data extraction, data analysis, writing of the manuscript, and had full access to and verified the existence of the raw data. MG conceptualised the study, participated in data extraction, data analysis, writing of the manuscript, had full access to and verified the existence of the raw data, and had the final responsibility to submit the manuscript for publication. EM participated in data extraction and had full access to the data. JPF critically reviewed the methodology and the final manuscript and had full access to the data. TGJdM, MMT, and MAB critically reviewed the final manuscript and had full access to the data. All authors had full access to all data. VS, MG, and JG verified the data in the study and MG had final responsibility for the decision to submit for publication and all authors agreed to submission.

Declaration of interests

JPF reports honoraria from United Pharmaceuticals paid to the hospital for lectures and educational activity. MAB reports honoraria from United Pharmaceuticals, Danone, Abbott, paid to the hospital for lectures and educational activity. All other authors declare no competing interests.

Data sharing

A prospectively published protocol for this manuscript is available online. Most data are shared in the online appendices. However, further data is available upon reasonable request.

For the **published protocol** see https://clok.uclan.ac.uk/52787/

Acknowledgments

We would like to extend our gratitude to Shiyao Liu, Aderonke Ajiboye, and Daniel Arruda Navarro Albuquerque for their contribution to data collection and title and abstract screening, and Yuhong Yuan for reviewing, updating and running the search strategy.

References

1 Korterink JJ, Diederen K, Benninga MA, Tabbers MM. Epidemiology of pediatric functional abdominal pain disorders: a meta-analysis. *PLoS One* 2015; **10:** e0126982.
2 Youssef NN, Murphy TG, Langseder AL, Rosh JR. Quality of life for children with functional abdominal pain: a comparison study of patients' and parents' perceptions. *Pediatrics* 2006; **117:** 54–59.
3 Varni JW, Lane MM, Burwinkle TM, et al. Health-related quality of life in pediatric patients with irritable bowel syndrome: a comparative analysis. *J Dev Behav Pediatr* 2006; **27:** 451–58.
4 Koppen IJ, Nurko S, Saps M, Di Lorenzo C, Benninga MA. The pediatric Rome IV criteria: what's new? *Expert Rev Gastroenterol Hepatol* 2017; **11:** 193–201.
5 de Bruijn CM, Rexwinkel R, Gordon M, Sinopoulou V, Benninga MA, Tabbers MM. Dietary interventions for functional abdominal pain disorders in children: a systematic review and meta-analysis. *Expert Rev Gastroenterol Hepatol* 2022; **16:** 359–71.
6 Gordon M, Sinopoulou V, Tabbers M, et al. Psychosocial interventions for the treatment of functional abdominal pain disorders in children: a systematic review and meta-analysis. *JAMA Pediatr* 2022; **176:** 560–68.
7 Wallace C, Gordon M, Sinopoulou V, Akobeng AK. Probiotics for management of functional abdominal pain disorders in children. *Cochrane Database Syst Rev* 2023; **2:** CD012849. https://doi.org/10.1002/14651858.CD012849.pub2.
8 Rexwinkel R, de Bruijn CMA, Gordon M, Benninga MA, Tabbers MM. Pharmacologic treatment in functional abdominal pain disorders in children: a systematic review. *Pediatrics* 2021; **147:** e2020042101.
9 Hoekman DR, Zeevenhooven J, van Etten-Jamaludin FS, et al. The placebo response in pediatric abdominal pain-related functional gastrointestinal disorders: a systematic review and meta-analysis. *J Pediatr* 2017; **182:** 155–63.

10 Gordon M, Benninga MA, Borlack R, et al. ESPGHAN and NASPGHAN 2023 protocol for paediatric FAPD treatment guidelines (standard operating procedure). *BMJ Paediatr Open* 2023; **7:** e002166.

11 Salanti G, Nikolakopoulou A, Efthimiou O, Mavridis D, Egger M, White IR. Introducing the treatment hierarchy question in network meta-analysis. *Am J Epidemiol* 2022; **191:** 930–38.

12 Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; **64:** 163–71.

13 Izcovich A, Chu DK, Mustafa RA, Guyatt G, Brignardello-Petersen R. A guide and pragmatic considerations for applying GRADE to network meta-analysis. *BMJ* 2023; **381:** e074495.

14 Schünemann HJ, Neumann I, Hultcrantz M, et al, and the GRADE Working Group. GRADE guidance 35: update on rating imprecision for assessing contextualized certainty of evidence and making decisions. *J Clin Epidemiol* 2022; **150:** 225–42.

15 Brignardello-Petersen R, Izcovich A, Rochwerg B, et al, and the GRADE working group. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. *BMJ* 2020; **371:** m3907.

16 Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (eds). Cochrane Handbook for Systematic Reviews of Interventions, version 6.5. https://training.cochrane.org/handbook/current/chapter-11 (accessed Jan 16, 2025).

17 Gordon M, Groen J, Sinopoulou V, Tabbers M, Moughey E, Franciosi J. Efficacy and safety of therapies for FAPD (FAP-NOS, IBS) in children: protocol for a network meta-analysis. https://clok.uclan.ac.uk/52787/ (accessed Jan 27, 2025).

18 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021; **372**: n71.

19 Thapar N, Benninga MA, Crowell MD, et al. Paediatric functional abdominal pain disorders. *Nat Rev Dis Primers* 2020; **6:** 89.

20 Higgins JPT, Altman DG, Gøtzsche PC, et al, and the Cochrane Bias Methods Group, and the Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; **343:** d5928.

21 Daly C, Welton NJ, Dias S, Answer S, Ades AE. Meta-analysis of continuous outcomes: guideline methodology document 2. 2021. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.bristol.ac.uk/media-library/sites/social-community-medicine/documents/mpes/gmd-2-continuous-jan2021.pdf (accessed Jan 16, 2025).

22 Rücker G, Petropoulou M, Schwarzer G. Network meta-analysis of multicomponent interventions. *Biom J* 2020; **62:** 808–21.

23 Gordon M. Maintaining remission in Crohn's disease post surgery: what can we learn from Cochrane? *Frontline Gastroenterol* 2024; **15:** 241–46.

24 Zeevenhooven J, Rexwinkel R, Van Berge Henegouwen VWA, et al. a core outcome set for clinical trials in pediatric functional abdominal pain disorders. *J Pediatr* 2020; **221:** 115–122.

25 Gordon M, Lakunina S, Sinopoulou V, Akobeng A. Minimum sample size estimates for trials in inflammatory bowel disease: a systematic review of a support resource. *World J Gastroenterol* 2021; **27:** 7572–81.

26 Iheozor-Ejiofor Z, Lakunina S, Gordon M, Akintelure D, Sinopoulou V, Akobeng A. Sample-size estimation is not reported in 24% of randomised controlled trials of inflammatory bowel disease: A systematic review. *United European Gastroenterol J* 2021; **9:** 47–53.

27 Gordon M, Khudr J, Sinopoulou V, Lakunina S, Rane A, Akobeng A. Quality of reporting inflammatory bowel disease randomised controlled trials: a systematic review. *BMJ Open Gastroenterol* 2024; **11:** e001337.

28 Sinopoulou V, Shah E, Gordon M, Tony-Jimmy TE. Primary author contact for systematic reviews of randomized controlled trials: a systematic review. *World J Methodol* 2024; **15:** 95559.

29 McLellan LF, Stapinski LA, Peters L. Pretreatment CBT-mindedness predicts CBT outcome. *Cognit Ther Res* 2019; **43:** 303–11.

30 Morgano GP, Mbuagbaw L, Santesso N, et al. Defining decision thresholds for judgments on health benefits and harms using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) frameworks: a protocol for a randomised methodological study (GRADE-THRESHOLD). *BMJ Open* 2022; **12:** e053246.

31 Chiocchia V, Nikolakopoulou A, Papakonstantinou T, Egger M, Salanti G. Agreement between ranking metrics in network meta-analysis: an empirical study. *BMJ Open* 2020; **10:** e037744.

32 Schulz KF, Altman DG, Moher D, and the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; **340:** c332.

33 Gordon M, Sinopoulou V, Mardare R, Abdulshafea M, Grafton-Clarke C, Vasiliou J. Patients' and health professionals' research priorities for chronic pain associated with inflammatory bowel disease: a co-produced sequential mixed methods Delphi consensus study. *BMJ Open Gastroenterol* 2024; **11:** e001483.