How fragile are Mediterranean diet interventions? A research-on-research study of randomised controlled trials

Maria G Grammatikopoulou, Meletios P Nigdelis, Xenophon Theodoridis, Konstantinos Gkiouras, Antigoni Tranidou, Theodora Papamitsou, Dimitrios P Bogdanos, Dimitrios G Goulis

ABSTRACT

Introduction The Mediterranean diet (MD) is a traditional regional dietary pattern and a healthy diet recommended for the primary and secondary prevention of various diseases and health conditions. Results from the higher level of primary evidence, namely randomised controlled trials (RCTs), are often used to produce dietary recommendations; however, the robustness of RCTs with MD interventions is unknown.

Methods A systematic search was conducted and all MD RCTs with dichotomous primary outcomes were extracted from PubMed. The fragility (FI) and the reverse fragility index (RFI) were calculated for the trials with significant and non-significant comparisons, respectively.

Results Out of 27 RCTs of parallel design, the majority failed to present a significant primary outcome, exhibiting an FI equal to 0. The median FI of the significant comparisons was 5, ranging between 1 and 39. More than half of the comparisons had an FI <5, indicating that the addition of 1-4 events to the treatment arm eliminated the statistical significance. For the comparisons with an FI=0, the RFI ranged between 1 and 39 (Median RFI: 7).

When the included RCTs were stratified according to masking, the use of a composite primary endpoint, sample size, outcome category, or dietary adherence assessment method, no differences were exhibited in the FI and RFI between groups, except for the RFI among different compliance assessment methods.

Conclusions In essence, the present study shows that even in the top tiers of evidence hierarchy, research on the MD may lack robustness, setting concerns for the formulation of nutrition recommendations.

BACKGROUND

Since Keys first presented a diet-mortality hypothesis explaining the Seven Countries study results in 1986, the Mediterranean diet (MD) has become a dietary pattern of particular interest. Research on the MD has spiralled, reputed for its health effects, spanning from ameliorated cardiovascular disease (CVD) factors, to improved pregnancy outcomes, ticking all the boxes in the quest for health attainment. For some, the MD is much more than a traditional regional dietary pattern, being regarded as the ‘unicorn’ of diet paradigms, with many clinical practice guidelines endorsing the adherence of the MD.

Apart from many ‘followers’ however, several scientists are also questioning the MD. Some are high-lightening the observational design of the Seven Countries study, while others are stressing the limitations of nutritional epidemiology in general, often incorporating selective reporting, inflated results, over-interpretation and skewed perspectives, with large flexibility in the performed analyses which can be based on questionnaires of low reproducibility.

Subsequently, research designs were improved to minimise bias, and the focus shifted to randomised controlled trials (RCTs), situated higher in the pyramid of evidence. The worm turned again when the biggest and most promising MD trial to date, the Prevención con Dieta Mediterránea (PREDIMED), raised concerns over randomisation bias, resulting in its reanalysis. Nutrition RCTs were once more in the spotlight, and scepticism was apparent, with researchers questioning the suitability of RCTs for nutrition research and the quality
of the trials. Most trials tend to report positive findings; however, statistical significance (P-value) does not ensure the robustness of an analysis and a pledge towards the use of more specific measures was made.

Today, clinical research continues to emphasise the P threshold of 0.05 when interpreting RCT results. For this, it is additionally important to evaluate the robustness of RCTs with MD interventions and attain an additional measure of the quality of MD RCTs. Two indexes have been proposed for the evaluation of an RCTs’ robustness; namely the fragility index (FI) and the reverse fragility index (RFI), for trials with significant or non-significant findings, respectively. Both indexes can only be calculated on studies with an RCT design and dichotomous primary outcomes.

To assess the robustness of RCTs with MD interventions, the present research-on-research study aimed to identify all RCTs with MD interventions and dichotomous primary outcomes, and calculate their FI or RFI, depending on the significance of the comparisons.

**METHODS**

**Research question and search strategy**

The present study used a systematic search strategy to answer the question “What is the fragility and reverse fragility index of RCTs assessing MD interventions?” The PICO of the study’s hypothesis was P: human population of any age group or health status, I: MD intervention, C: any comparison other than the MD, a sham diet, other diet or no intervention, O: any dichotomous primary outcome (table 1). To answer this research question, the focus was set on all RCTs examining MD interventions, irrespective of their other characteristics. Similar studies examining the FI/RFI in broad research areas are common in the literature.

The protocol of the study was published at the Center for Open Science https://osf.io/mnx2c/. A systematic search was conducted on PubMed from inception until 31 August 2019, using the keyword (Mediterranean diet) and the PubMed filter for clinical trials.

**Inclusion and exclusion criteria**

As the concept of fragility is only applicable to RCTs, only studies with an RCT design were considered eligible. In parallel, we searched for trials with dichotomous primary outcomes, as the FI and RFI cannot be calculated in trials with continuous outcomes. Secondary outcomes were not of concern as they are not accounted for when estimating the sample size required for an RCT and should not be used to assess a trial’s robustness. All RCTs with MD interventions were assessed for eligibility, despite other possible heterogeneities, as the research question focused on the FI and RFI of MD interventions in general and not in MD RCTs with more homogenous outcomes/samples/designs.

The criteria for inclusion in the present analyses involved (1) RCTs performed on humans, (2) of any age group, (3) irrespectively of any medical diagnosis or health condition, (4) applying MD interventions, (5) compared with no intervention, control diet, or to dietary patterns other than the MD, (6) assessing any dichotomous primary outcome and (7) published in any language.

On the other hand, criteria for exclusion involved trials (1) lacking randomisation, (2) performed on animals, (3) with continuous primary outcomes or (4) with dichotomous secondary outcomes, (5) comparing MD interventions to control diets based on the MD, (6) not including an MD intervention, (7) not reporting the number of events and the sample size in each arm, making it impossible to calculate 2×2 frequency tables, (8) failing to report adequate data to calculate persons–years, (9) trial protocols without results and (10) research performed on animals.

**Data extraction**

Two researchers (MGG and XT) independently extracted data from the selected RCTs, aided by an additional pair of reviewers (MPN and KG) when deemed necessary. Extracted data involved details regarding the study design, the level of masking (open label/single/double), sample size, protocol registration details, study name/acronym, interventions and comparators, the primary outcomes, the event rates in each arm, the geographical origin of the trial, the randomisation methods used, the level of prevention (primary/secondary) and the methods used to assess intervention adherence. As far as time-to-event outcomes are concerned, extracted data involved the total number of events in each arm over the entire follow-up period of each trial.

**Risk of bias**

The risk of bias (RoB) of the selected RCTs was evaluated using the Cochrane RoB V.2.0 tool by two independent researchers (MPN and KG). Disagreements were resolved via discussion and whenever needed, through the intervention of more experienced researchers (DGG, MGG and DPB).

**Calculation of the FI and RFI**

The FI was developed as a measure of RCT robustness. It describes the minimum number of patients within the group with the fewest event count needed to change from...
a non-event to an event, to transform a significant result to a non-significant one. It is considered as the measurement of the event count, on which the statistical significance depends.

For the current analysis, two researchers (XT and MGG) calculated the FI of each RCT, according to Walsh et al. In further detail, after extracting the number of events and non-events for each trial arm in 2×2 tables, the additional number of events required to be added in the group with the smaller number of events to make the p value of the Fisher’s exact test ≥0.05 was calculated.

An FI equal to zero describes a highly fragile RCT, as zero participants are required to change from a non-event to an event to reverse a significant finding to a non-significant one.

On the other hand, in non-significant comparisons (with an FI equal to 0), the RFI was calculated. This was performed via the subtraction of events from the arm with the fewer events, while simultaneously adding non-events to the same arm, keeping the number of total participants constant, until the Fisher exact test two-sided P-value became <0.05. Lower RFIs indicate reduced statistical robustness and increased vulnerability to change from statistical non-significance to significance, with only a minimum number of events. At the moment, there is no recognised cut-off for categorising either the FI or the RFI.

For the current analyses, 2×2 tables were created in Microsoft Excel and the Fischer’s exact test was used to calculate and verify the FIs and RFIs of the included trials. For one trial, the reported sample and events in each group were used to calculate the FI, and for another, the incidence and the total number of participants allocated in each group were applied in the FI calculations. When more than two interventions were included in one trial, like in the PREDIMED, each arm was compared with the control diet independently, and the FI or RFI was calculated accordingly, for each paired comparison. When the primary outcome was not reported, the first result presented in the abstract was considered as the primary outcome. In RCTs reporting more than one dichotomous primary outcomes, the FI of all three endpoints was calculated accordingly.

### Statistical analyses

As the research question was ‘broad’, incorporating all RCTs with MD interventions, an effort to assess differences in RCTs with different characteristics was also performed. Three researchers (KG, MPN and MGG) stratified the selected trials according to blinding, outcome category, sample size, the use of a composite outcome (yes/no) and the method used to assess compliance to the assigned dietary scheme. These categories were used to detect differences in the FI and the RFI between RCTs with different design characteristics and outcomes. As most data did not follow the normal distribution hypothesis, results were presented as medians with their respective IQRs. Group differences were assessed with the Mann-Whitney U test (for comparisons involving two groups) and the Kruskal-Wallis test (for comparisons involving more than two groups). For these analyses, the

---

**Figure 1** Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart of the selection of the studies. MD, Mediterranean diet; RCT, randomised controlled trial.
Jamovi project (V.0.9.5.16) was used. Significance was set at 0.05, unless otherwise specified.

RESULTS
Search results and RCT characteristics
The detailed process of the selection of RCTs fulfilling the study’s criteria is illustrated in figure 1. Published protocols of RCTs lacking the reporting of results, published studies with design issues (cross-sectional, qualitative, or post-hoc analyses), RCTs without dichotomous primary outcomes, and trials lacking a MD intervention, or those with a concomitant MD comparator arm were excluded from the records. A total of 35 distinct publications18 19 31–63 of parallel interventions were identified meeting the predefined criteria (table 2), with those having an original publication and an erratum being counted as one record (five cases in total).18 19 37–39 44 46–48 63 Multiple publications deriving from the same trials, using the same sample size and outcomes, were also counted as one record (three cases in total).31–33 35 36 53 This resulted in 27 distinct RCTs in total, fulfilling the study’s criteria and being included in the present analyses.

The majority of RCTs were performed in Spain,18 19 35 37–44 46–48 50 52–56 59–61 63 four originated from France,34–36 62 two took place in the UK31 58 and in Italy,31 32 and single trials were performed in Australia,15 Israel,57 and India.49 Most publications belonged to the PREDIMED or PREDIMED Reus trials,18 19 37–39 44 46–48 63–65 and few referred to the Lyon Heart Study.34–36 62 Two records involved the St. Carlos gestational diabetes mellitus (GDM) prevention RCT,39 66 and others were produced from the Effect of Simple, Targeted Diet in Pregnant Women With Metabolic Risk Factors on Pregnancy Outcomes (ESTEEM),58 The Heart Institute of Spokane Diet Intervention and Evaluation Trial,51 Pre Frail 80,41 Indo-MD Heart Study49 or other trials.31 32 42 45 57 (table 2). The sample size ranged from 5645 to 740344 participants.

Given the nature of the intervention (diet), most RCTs were of single-blind masking, and the remaining were open labelled. Regarding the PREDIMED trial, the single-blind masking was disputed by some researchers and further verifications were published by the investigators to support the issue.

Intervention and outcomes
For one trial,41 it was difficult to discern the exact primary outcome. For this specific RCT,41 the first result presented in the abstract (reversion to robustness) was considered as the primary outcome. Accordingly, given that the Consolidated Standards of Reporting Trials (CONSORT)64 guidelines were produced fairly recently, few—mainly older—trials did not have a preregistered protocol, although some had preceding publications detailing the protocol.

The PREDIMED RCTs18 19 33 37–40 43 44 46–48 50 52–56 61 63 evaluated the efficacy of two MD interventions, one with extra-virgin olive oil (EVOO) and one with nuts, in a great variety of health outcomes. In further detail, included PREDIMED RCTs involved the prevention the development of diabetic retinopathy and nephropathy,37 38 CVD,18 19 incidence and reversion of the metabolic syndrome,55 61 liver steatosis,52 depression,56 osteoporosis-related fractures,33 53 peripheral artery disease,54 the occurrence of cataract surgery,40 as well as the incidence of type 2 diabetes mellitus (T2DM),16–58 atrial fibrillation,43 breast cancer,49 and heart failure.48 Among the remaining trials, the majority34–36 42 49 51 57 62 investigated the effects of the MD on CVD risk factors. The St. Carlos GDM prevention59 60 and the ESTEEM58 trials used a MD with EVOO and pistachios to investigate maternal and fetal outcomes. The Pre Frail 80 and Properzi45 trials applied the MD to evaluate frailty50 and non-alcoholic fatty liver (NAFLD) parameters, respectively.

Compliance to the dietary interventions was assessed by the majority of trials using the MEDAS questionnaire,66 food frequency questionnaires (FFQs)34–36 57–58 including the ESTEEM-O2,67 previous day 24 hour diet recall57 and diet diaries,31 32 45 49 51 57 62 diet history,45 diet ‘surveys’ and other questionnaires.42 58 In addition, biomarkers indicative of increased MD adherence were selectively assessed, including urinary hydroxytyrosol concentrations and plasma γ-linolenic acid proportions.

When adherence to the control diet differed from that of the intervention group, either a 9-item dietary screener was used, or compliance assessment was not reported in the procedures at all.

Risk of bias
A summary of the RoB of the included RCTs is presented in figure 2. For some of the PREDIMED RCTs,18 19 37–46 the deviations from the randomisation protocol were considered when assessing the domains of random-sequence generation and allocation-sequence concealment. Many of the PREDIMED RCTs18 19 37–39 44 46–48 63 published errata and reanalyses of their datasets, excluding participants who had deviated from the randomisation protocol; for these, the allocation sequence concealment was considered as adequate, without, however, altering the random-sequence generation domain of the RoB tool, which remained biased. Furthermore, the use of different tools to assess compliance between intervention and controls was also accounted for when assessing the RoB, as it confuted the single-blind masking.

According to the RoB (figure 2), the majority of RCTs exhibited either unclear, or high overall bias.18 19 33 35 36 38–42 46 47 50 52 54–57 59–61 The fewest concerns were raised with regard to missing outcome data. Among all included RCTs, the ESTEEM58 demonstrated the lowest bias throughout the examined RoB domains.

FI and RFI of the included RCTs
Table 3 details the FI and RFI of all included RCTs. The majority of comparisons33 37–40 43–48 50–52 55 57 58 61 failed to provide a significant result between MD intervention and comparator arms, exhibiting an FI equal to 0. On the other hand, the FI of significant comparisons ranged between 152 and 39.66 The median FI of the RCTs, excluding those with non-significant comparisons, was 5. More than half of the comparisons had an FI <5, indicating that the addition of 1–4 events to the opposite treatment arm
Table 2  Characteristics of the included trials

<table>
<thead>
<tr>
<th>First author</th>
<th>Origin</th>
<th>Trial name</th>
<th>Design, masking</th>
<th>Registry</th>
<th>Participants*</th>
<th>Randomisation</th>
<th>Prevention tier</th>
<th>Primary outcome</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al Wattar et al.</td>
<td>UK</td>
<td>ESTEEM</td>
<td>Pragmatic, Parallel, Single blind</td>
<td>NCT02 218931</td>
<td>N=1252 inner-city pregnant women with metabolic risk factors (obesity, hypertension, or hypertriglyceridaemia)</td>
<td>1:1 ratio via a password-protected on-line data management system</td>
<td>Primary</td>
<td>Maternal composite outcome, † offspring composite outcome‡</td>
<td>MD high in nuts, EVOO, fruits, vegetables, non-refined grains and legumes, moderate-to-high fish, low-to-moderate poultry and dairy, low intake of red/processed meat, avoidance of sugar, fast food and food rich in animal fat</td>
<td>Usual care (dietary advice)</td>
<td>FFQ and ESTEEM Q</td>
</tr>
<tr>
<td>Assaf-Balut et al.</td>
<td>Spain</td>
<td>St. Carlos GDM prevention study</td>
<td>Parallel, Open label</td>
<td>ISRCTN8 4389045</td>
<td>N=1000 normoglycaemic (&lt;92 mg/dL) pregnant women at 8±12 gestational wk</td>
<td>Stratified with permuted block randomisation, by age, prepregnant BMI, ethnicity, parity, in a 1:1 ratio and 4-6 blocks</td>
<td>Primary</td>
<td>GDM incidence</td>
<td>MD supplemented with EVOO and pistachios (&gt;40 mL of EVOO and 25–30 g of pistachios each day)</td>
<td>Standard diet with limited fat intake</td>
<td>MEDAS, DNT FFQ, serum γ-tocopherol</td>
</tr>
<tr>
<td>Assaf-Balut et al.</td>
<td>Spain</td>
<td>St. Carlos GDM prevention study</td>
<td>Parallel, Open label</td>
<td>ISRCTN8 4389045</td>
<td>N=697 normoglycaemic (&lt;92 mg/dL) pregnant women at 8±12 gestational wk</td>
<td>Stratified with permuted block randomisation, by age, prepregnant BMI, ethnicity, parity, in a 1:1 ratio and 4-6 blocks</td>
<td>Primary</td>
<td>Composite maternofetal outcome‡</td>
<td>MD supplemented with EVOO and pistachios (&gt;40 mL of EVOO, 25–30 g of pistachios every day)</td>
<td>Standard diet with limited fat intake</td>
<td>MEDAS, DNT FFQ, urine HXT, serum γ-tocopherol</td>
</tr>
<tr>
<td>Babio et al.</td>
<td>Spain</td>
<td>PREDIMED</td>
<td>Parallel, single blind</td>
<td>ISRCTN3 5739639</td>
<td>N=5801 men/women (55-80 years) with T2DM and/or ≥3 CVD risk factors†</td>
<td>1:1:1 PC-generated randomisation table*</td>
<td>Primary</td>
<td>MetS</td>
<td>1. MD with EVOO (1 L/wk for the participants and families) &lt;br&gt;2. MD with mixed nuts (30 g: 15 g walnuts, 7.5 g hazelnuts, 7.5 g almonds)</td>
<td>Control diet (advice to reduce dietary fat)</td>
<td>MEDAS,9-item dietary screener (per arm)</td>
</tr>
<tr>
<td>de Lorgeri et al.</td>
<td>France</td>
<td>Lyon Diet Heart Study</td>
<td>Parallel, single-blind</td>
<td>NR</td>
<td>N=423 consecutive patients who survived a first MI at 6 months of enrolment</td>
<td>NR</td>
<td>Secondary</td>
<td>Composite outcome† †</td>
<td>MD: more bread, root, green vegetables and fish, less meat (beef/lamb/pork replaced by poultry), no day without fruit. Butter/cream replaced by canola margarine</td>
<td>Prudent Western-type diet</td>
<td>Diet ‘survey’ (MD group), plasma FA</td>
</tr>
<tr>
<td>de Lorgeri et al.</td>
<td>France</td>
<td>Lyon Diet Heart Study</td>
<td>Parallel, single-blind</td>
<td>NR</td>
<td>N=605 consecutive patients who survived a first MI within 6 months of enrolment</td>
<td>NR</td>
<td>Secondary</td>
<td>CV mortality, non-fatal MI</td>
<td>MD: more bread, root, green vegetables and fish, less meat (beef/lamb/pork replaced by poultry), no day without fruit. Butter/cream replaced by canola margarine</td>
<td>Prudent Western-type diet</td>
<td>24 hours recall and FFQ</td>
</tr>
<tr>
<td>de Lorgeri et al.</td>
<td>France</td>
<td>Lyon Diet Heart Study</td>
<td>Parallel, single-blind</td>
<td>NR</td>
<td>N=605 consecutive patients who survived a first MI within 6 months of enrolment</td>
<td>NR</td>
<td>Secondary</td>
<td>CV mortality, non-fatal MI</td>
<td>MD: more bread, root, green vegetables and fish, less meat (beef/lamb/pork replaced by poultry), no day without fruit. Butter/cream replaced by canola margarine</td>
<td>Prudent Western-type diet</td>
<td>24 hours recall and FFQ</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Origin</th>
<th>Trial name</th>
<th>Design, masking</th>
<th>Registry</th>
<th>Publication (journal, year)</th>
<th>Participants*</th>
<th>Randomisation</th>
<th>Prevention tier</th>
<th>Primary outcome</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Compliance</th>
</tr>
</thead>
</table>
| Díaz-López [27] [30] | Spain     | PREDIMED                        | Parallel, Single blind | ISRCTN3 5739639  | Diabetes Care, 2015; Rev Esp Cardiol, 2019 | N=3614 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors| 1:1:1 PC-generated randomisation table** | Primary          | New-onset of diabetic retinopathy, nephropathy | 1. MD with EVOO (1 L/wk for the participant and their families)  
2. MD with mixed nuts (30 g/d: 15 g walnuts, 7.5 g hazelnuts, 7.5 g almonds) | Control diet (advice to reduce dietary fat) | Urine HXT and plasma ALA proportions |
| Esposito [11][32]   | Italy     | –                               | Parallel, open-label | NCT00 725297     | Diabetes Care, 2014; Ann Intern Med, 2009 | N=215 men/women (50–75 years) with newly diagnosed T2DM | ‘Simple’ randomisation; PC-generated random sequence | Secondary         | Initiation of T2DM medication                  | LOMD                                                                             | LFD                         | Diet diaries |
| Estruch [18][19]    | Spain     | PREDIMED                        | Parallel, Single-blind | ISRCTN3 5739639  | NEJM, 2018                  | N=7447 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors | 1:1:1 sealed envelopes (pilot phase) and PC-generated random number | Primary          | Composite CV events†‡                         | 1. MD with EVOO (1 L/wk for the participant and their families)  
2. MD with mixed nuts (30 g/d: 15 g walnuts, 7.5 g hazelnuts, 7.5 g almonds) | Control diet (advice to reduce dietary fat) | MEDAS/8-item dietary screener, urine HXT, plasma ALA |
| García-Gavilán [33][53] | Spain     | PREDIMED Reus                    | Parallel, Single-blind | ISRCTN3 5739639  | Clin Nutr, 2018; Am J Clin Nutr, 2018 | N=870 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors | 1:1:1 PC-generated randomisation table** | Primary          | Osteoporotic fractures                        | 1. MD with EVOO (1 L/wk for the participant and their families)  
2. MD with mixed nuts (30 g/d: 15 g walnuts, 7.5 g hazelnuts, 7.5 g almonds) | Control diet (advice to reduce dietary fat) | MEDAS          |
| García-Layana [40]  | Spain     | –                               | Parallel, Single-blind | ISRCTN3 5739639  | Nutrients, 2017              | N=5802 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors | 1:1:1 PC-generated randomisation table** | Primary          | Occurrence of cataract surgery                 | 1. MD with EVOO (1 L/wk for the participant and their families)  
2. MD with mixed nuts (30 g/d: 15 g walnuts, 7.5 g hazelnuts, 7.5 g almonds) | Control diet (advice to reduce dietary fat) | MEDAS, urine HXT and plasma ALA ratio |
| GenéHuguet [41]     | Spain     | Pre Frail 80 Study              | Parallel, open-label | NR                | J Nutr Health Aging, 2018    | N=200 non-institutionalised men/women (≥80 years) fulfilling 1/2 of the Fried frailty criteria | Randomised list NOD | Secondary | Reversion to robustness                      | MD                                                                            | Standard treatment | MEDAS          |
| Greenberg [23]      | Israel    | DIRECT                          | Parallel, NR     | NCT001 60108     | J Am Coll Nutr, 2009         | N=322 men/women (40–65 years) with BMI ≥27 kg/m²/T2DM/CHD | Based on sex, age, BMI, history of CHD/T2DM and statins use | 5% BW loss       | Hypocaloric §§ MD based on Willett. Fat: 40% (mainly olive oil and nuts) | (1) Atkins-based LCD. CHO 20 g/d (first 2 months), ≤100 g/d thereafter. Unlimited EL, protein, fat.  
(2) Hypocaloric §§ LFD based on the AHA. Fat 20%–30% (SFA 7%–10%, 200–300 mg cholesterol) | 127-item FFQ with 3 portion size pictures for 17 selected items |  |
| Marcos-Forniol [42] | Spain     | –                               | Parallel, open-label | ISRCTN1 7382091  | Eur J Prev Cardiol, 2018     | N=127 consecutive patients (≥70 years) with acute coronary syndrome §§ §§ §§ | 1:1 PC allocation with random block sizes of 2 | Secondary | Optimal CVD risk factor control**            | MD                                                                            | Standard care 9-item MD score |  |

Table 2 Continued
### Table 2

<table>
<thead>
<tr>
<th>First author, Origin</th>
<th>Trial name</th>
<th>Design, masking</th>
<th>Registry</th>
<th>Publication (journal, year)</th>
<th>Participants</th>
<th>Randomisation</th>
<th>Prevention tier</th>
<th>Primary outcome</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-González et al.</td>
<td>Spain</td>
<td>PREDIMED</td>
<td>Parallel, open-label</td>
<td>ISRCTN3 5739639</td>
<td>Circulation, 2014</td>
<td>N=7447 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors</td>
<td>1:1:1 PC-generated randomisation table</td>
<td>Primary</td>
<td>Incidence of atrial fibrillation</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
<tr>
<td>Papadaki et al.</td>
<td>Spain</td>
<td>PREDIMED</td>
<td>Parallel, open-label</td>
<td>ISRCTN3 5739639</td>
<td>Eur J Heart Fail, 2017</td>
<td>N=7403 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors</td>
<td>1:1:1 PC-generated randomisation table</td>
<td>Primary</td>
<td>Heart failure incidence</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
<tr>
<td>Papadaki et al.</td>
<td>Spain</td>
<td>PREDIMED</td>
<td>Parallel, open-label</td>
<td>ISRCTN3 5739639</td>
<td>J Nutr, 2019</td>
<td>N=109 consecutive patients, men/women (55-80 years) with T2DM and/or ≥3 CVD risk factors</td>
<td>1:1:1 PC-generated randomisation table</td>
<td>Primary</td>
<td>Steatosis diagnosis</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
<tr>
<td>Properzi et al.</td>
<td>Australia</td>
<td>–</td>
<td>Parallel, single-blind</td>
<td>ACTRN1 2612000 841875</td>
<td>Hepatology, 2018</td>
<td>N=56 adult patients with NAFLD</td>
<td>In a 1:1 fashion using randomly selected envelope-concealed allocations in blocks of 4</td>
<td>Primary</td>
<td>NAFLD resolution</td>
<td>MD based on foods consumed in traditional Cretan diet, altered to allow for standardisation of protein intake with the control diet. CHO: 40%, fat: 35–40% (&lt;10% SFA), protein: &lt;20%</td>
<td>LFD (based on NHMRC and AHA). CHO: 50%, fat: 30%, SFA &lt;10%, protein: 20%</td>
</tr>
<tr>
<td>Ruiz-Canela et al.</td>
<td>Spain</td>
<td>PREDIMED</td>
<td>Parallel, open-label</td>
<td>ISRCTN3 5739639</td>
<td>JAMA, 2014</td>
<td>N=491 men/women (55–80 years) PAD-free and CVD-free but with T2DM and/or ≥3 CVD risk factors</td>
<td>1:1:1 ratio</td>
<td>Primary</td>
<td>New symptomatic PAD events</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
<tr>
<td>Salas-Salvadó et al.</td>
<td>Spain</td>
<td>PREDIMED</td>
<td>Parallel, open-label</td>
<td>ISRCTN3 5739639</td>
<td>Arch Intern Med, 2008</td>
<td>N=382 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors</td>
<td>NR but based on the PREDIMED protocol 1:1:1 PC-generated randomisation table</td>
<td>Primary and secondary</td>
<td>MetS reversal rate and incidence</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
<tr>
<td>Salas-Salvadó et al.</td>
<td>Spain</td>
<td>PREDIMED Reus</td>
<td>Parallel, open-label</td>
<td>ISRCTN3 5739639</td>
<td>Diabetes Care, 2011; 2018</td>
<td>N=418 non-diabetic, CVD-free men/women with ≥3 CVD risk factors</td>
<td>1:1:1 PC-generated randomisation table</td>
<td>Primary</td>
<td>Diabetes incidence</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
</tbody>
</table>

Continued
## Table 2 continued

<table>
<thead>
<tr>
<th>First author</th>
<th>Origin</th>
<th>Trial name</th>
<th>Design, masking</th>
<th>Registry</th>
<th>Participants*</th>
<th>Randomisation</th>
<th>Prevention tier</th>
<th>Primary outcome</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salas-Salvadó</td>
<td>Spain</td>
<td>PREMIDED</td>
<td>Parallel, open-label</td>
<td>ISRCTN3 573939</td>
<td>Ann Intern Med, 2014; 2018</td>
<td>N=3541 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors</td>
<td>1:1:1 PC-generated randomisation table</td>
<td>Primary</td>
<td>New-onset of diabetes</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
<tr>
<td>Sánchez-Villegas</td>
<td>Spain</td>
<td>PREMIDED</td>
<td>Parallel, open-label</td>
<td>ISRCTN3 573939</td>
<td>BMC Medicine, 2013</td>
<td>N=3923 men/women (55–80 years) with T2DM and/or ≥3 CVD risk factors</td>
<td>1:1:1 PC-generated randomisation table</td>
<td>Primary</td>
<td>Depression</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
<tr>
<td>Singh</td>
<td>India</td>
<td>Indo-MD Heart Study</td>
<td>Parallel, single-blind</td>
<td>NR</td>
<td>Lancet, 2002</td>
<td>N=1000 men/women (20–75 years) with hypercholesterolemia, hypertension, DM, angina pectoris, previous MI</td>
<td>By selection of a card from a pile of an equal number of cards for each group</td>
<td>Secondary</td>
<td>Total cardiac events</td>
<td>NCEP prudent diet (fat: 30%, SFA &lt;10%, cholesterol &lt;300 mg/d), &gt;400–500 g of fruit, vegetable and nuts, &gt;400–500 g whole-grains, legumes, rice, maize, wheat, 34 serv of mustard seed/soybean oil</td>
<td>NCEP prudent diet</td>
</tr>
<tr>
<td>Toledo</td>
<td>Spain</td>
<td>PREMIDED</td>
<td>Parallel, single-blind</td>
<td>ISRCTN3 573939</td>
<td>JAMA Intern Med, 2015</td>
<td>N=4282 women (60–80 years) with T2DM and/or ≥3 CVD risk factors</td>
<td>1:1:1 PC-generated randomisation table</td>
<td>Primary</td>
<td>Breast cancer incidence</td>
<td>1. MD with EVOO (1 L/wk for the participant and their families)</td>
<td>Control diet (advice to reduce dietary fat)</td>
</tr>
<tr>
<td>Tuttle</td>
<td>USA</td>
<td>THIS-DIET</td>
<td>Parallel, open-label (blind PI)</td>
<td>–</td>
<td>Am J Cardiol, 2008</td>
<td>N=101 MI survivors</td>
<td>Sealed envelopes with the allocation sequence, prepared by a PI, placed in a locked drawer</td>
<td>Primary</td>
<td>Free survival††</td>
<td>MD with fat: 30–40% (&lt;7% SFA), CHO: 50%, protein: 10–20%</td>
<td>LFD (AHA step I), Fat:&lt;30%, protein:10–20%, SFA&lt;7%, CHO:55–60%</td>
</tr>
</tbody>
</table>

*Number of initially randomised participants or in the secondary analyses of the Prevención con Dieta Mediterranea (PREMIDED) study, the number of initially randomised participants as stated in the respective papers.
†Maternal composite outcome: gestational diabetes mellitus (GDM) or preeclampsia. **34
‡Of offspring composite outcome: stillbirth, small-for-gestational age (SGA), or admission to neonatal care unit. §4
§3Concerns regarding randomisation rose post publication. ¶5
*Emergency cesarean section, perinatal trauma, pregnancy-induced hypertension and preeclampsia, prematurity, large-for-gestational age, and SGA.
**Smoking, hypertension, elevated low-density lipoprotein (LDL) level, low high-density lipoprotein level, overweight/obesity, or family history of premature coronary heart disease (CHD).
††3Cardiac death and non-fatal myocardial infarction (MI).
‡‡3MI, stroke, or death from cardiovascular (CV) causes.
§§3Hypertension, 1200–1500 kcal/day for women and 1500–1800 kcal/day for men.
¶¶3Elevation MB, non-ST elevation MI and unstable angina.
**Achievement of 6risk factor goals: blood pressure <140/90 mm Hg, LDL <2.6 mmol/L, smoking cessation, body mass index (BMI) <25 kg/m², physical activity of moderate intensity >30 min/day, 3 days/week (≥6 MET h/wk) and HbA₁c <7% in patients with diabetes.
††3Composite of all-cause and cardiac deaths, MI, hospital admissions for heart failure, unstable angina pectoris or stroke.
‡‡3MI, or death from cardiovascular (CV) causes.
AHA, American Heart Association; ALA, α-linolenic acid; BM, body weight; CHO, carbohydrate; CVD, cardiovascular disease; DIRECT, dietary intervention randomised controlled trial; DM, diabetes mellitus; DNCT, diabetes nutrition and complications trial; E1, energy intake; ESTEEM, Effect of Simple, Targeted Diet in Pregnant Women With Metabolic Risk Factors on Pregnancy Outcomes; EVOO, extra-virgin olive oil; FA, fatty acid; FFQ, food frequency questionnaire; HX, hydroxytyrosol; LCD, low-carbohydrate diet; LDM, low-fat diet; MD, Mediterranean diet; MEDAS, Mediterranean Diet Adherence Screener; MET, metabolic equivalents; MetS, metabolic syndrome; NHMRC, National Health and Medical Research Council; NOD, non-identified; NHL, not reported; PAD, peripheral artery disease; PC, personal computer; PI, principle investigator; ESTEEM Q, ESTEEM questionnaire; SFA, saturated fatty acids; ST, sinus tachycardia; T2DM, type 2 diabetes mellitus; THIS-DIET, The Heart Institute of Spokane Diet Intervention and Evaluation Trial.
eliminated the statistical significance of the RCTs. The most robust results (FI >15) involved publications of the St. Carlos GDM trial, the Indo-MD Heart Study, the Lyon Heart Study, and the PREDIMED comparison between MD + EVOO versus control diet, published by Babio et al.

For the comparisons with an FI=0, the RFI was calculated, ranging between 1.37–39.45 and 29.40 (median RFI: 7). Six out of 23 comparisons had RFI <5 (median: 4), indicating that the change of 1–4 non-events to events reverses the respective comparisons to statistically significant ones.

Categorisation of the FI and RFI according to study characteristics

Table 4 details the FI and RFI categorisation according to the RCT design, the number of participants, and the primary outcome. When masking was accounted for, no differences were noted in the FI or RFI between trials of different allocation masking.

Primary outcomes of the trials were categorised as perinatal, those related to diabetes mellitus or metabolic syndrome, cardiovascular, NAFLD-outcomes, or other (first incidence of breast cancer, cataract surgery, osteoporotic fractures, return to robustness or depression). This allocation failed to induce differences in the FI and RFI between different outcome categories. Similarly, allocation of the trials to those with composite primary endpoints against all others failed to show differences in the FI and RFI between the two groups.

Again, when sample size and methods used to assess dietary compliance between trials were used to allocate the RCTs, no differences were observed in the FI and RFI, with the exception of the RFI among distinct compliance-methods groups (p<0.035).

DISCUSSION

The present study revealed that most individual comparisons of RCTs with MD interventions and dichotomous primary outcomes as endpoints fail to demonstrate significant results. In parallel, those with comparisons yielding significant findings appear fragile, with a small number of events needed to change the result from significant to non-significant. Subsequently, the number of robust RCTs investigating MD interventions appears to be limited.

Among the reviewed trials, the St. Carlos GDM and the PREDIMED RCT conducted by Babio exhibited the highest FIs, indicating that nutrition RCTs can be robust. Both of these trials exhibited high RoB in several RoB domains, suggesting that robustness does not necessarily coincide with low RoB. In the St. Carlos GDM study, the reported event rate was high, corresponding to 27.8% and 25.8% of the intervention and control groups, respectively, whereas the Babio PREDIMED trial did not exhibit a similar high rate of events (1.5% of the total participants in the intervention arm and 3.9% of those allocated in the control group, respectively).
<table>
<thead>
<tr>
<th>First author</th>
<th>Outcome</th>
<th>Intervention</th>
<th>Events in intervention arm (n)</th>
<th>Events in control arm (n)</th>
<th>P-value (Intervention vs control)</th>
<th>Fragility Index (FI)</th>
<th>Reverse Fragility Index (RFI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al Wattar</td>
<td>Maternal composite outcome*</td>
<td>MD 58</td>
<td>486</td>
<td>111</td>
<td>0.04</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Assaf-Balut</td>
<td>Offspring composite outcome†</td>
<td>MD 531</td>
<td>434</td>
<td>92</td>
<td>0.14</td>
<td>4</td>
<td>39</td>
</tr>
<tr>
<td>Babio</td>
<td>GDM incidence</td>
<td>MD 1885</td>
<td>360</td>
<td>219</td>
<td>32</td>
<td>0.0001</td>
<td>39</td>
</tr>
<tr>
<td>Assaf-Balut</td>
<td>Composite maternal-fetal outcome‡</td>
<td>MD 1885</td>
<td>360</td>
<td>219</td>
<td>32</td>
<td>0.0001</td>
<td>39</td>
</tr>
<tr>
<td>Garcia-Layana</td>
<td>Osteoporotic fractures</td>
<td>MD 11728</td>
<td>291</td>
<td>289</td>
<td>37</td>
<td>0.016</td>
<td>13</td>
</tr>
<tr>
<td>Estruch</td>
<td>Incidence of MetS</td>
<td>MD 11829</td>
<td>10365</td>
<td>88</td>
<td>83</td>
<td>0.024</td>
<td>5</td>
</tr>
<tr>
<td>Gené Huguet</td>
<td>Cataract surgery incidence</td>
<td>MD 11728</td>
<td>10365</td>
<td>88</td>
<td>83</td>
<td>0.024</td>
<td>5</td>
</tr>
<tr>
<td>Marcos-Forniol</td>
<td>Optimal CV risk factor control‡‡</td>
<td>MD 10279</td>
<td>9102</td>
<td>72</td>
<td>8</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>

Continued...
**Table 3 Continued**

<table>
<thead>
<tr>
<th>First author</th>
<th>Outcome</th>
<th>Intervention</th>
<th>Events in intervention arm (n)</th>
<th>Events in Control arm (n)</th>
<th>P-value (Intervention vs control)</th>
<th>Fragility Index (FI)</th>
<th>Reverse Fragility Index (RFI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pintó⁵⁶</td>
<td>Presence of steatosis</td>
<td>MD + EVOO</td>
<td>34 3 30 10</td>
<td>0.027</td>
<td>1</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Pintó⁵⁶</td>
<td>Presence of steatosis</td>
<td>MD + Nuts</td>
<td>36 12 30 10</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Properzi⁶⁵</td>
<td>NAFLD resolution</td>
<td>MD</td>
<td>26 3 25 9</td>
<td>0.046§§</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Ruiz-Canela⁵⁴</td>
<td>New symptomatic PAD cases</td>
<td>MD + EVOO</td>
<td>2539 18 2444 45</td>
<td>&lt;0.001</td>
<td>12</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Ruiz-Canela⁵⁴</td>
<td>New symptomatic PAD cases</td>
<td>MD + Nuts</td>
<td>2452 26 2444 45</td>
<td>0.023</td>
<td>3</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Salas-Salvadó⁵⁵</td>
<td>Reversion of MetS</td>
<td>MD + EVOO</td>
<td>252 53 250 41</td>
<td>0.208</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Salas-Salvadó⁵⁵</td>
<td>Reversion of MetS</td>
<td>MD + Nuts</td>
<td>249 63 250 41</td>
<td>0.015</td>
<td>4</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Salas-Salvadó⁵⁵</td>
<td>New MetS Incidence</td>
<td>MD + EVOO</td>
<td>157 36 154 36</td>
<td>1</td>
<td>0</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Salas-Salvadó⁵⁵</td>
<td>New MetS Incidence</td>
<td>MD + Nuts</td>
<td>162 29 154 36</td>
<td>0.266</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Salas-Salvadó⁵⁵</td>
<td>Incidence of T2DM</td>
<td>MD + EVOO</td>
<td>570 14 515 24</td>
<td>0.068</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Salas-Salvadó⁵⁵</td>
<td>Incidence of T2DM</td>
<td>MD + Nuts</td>
<td>598 16 515 24</td>
<td>0.105</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Salas-Salvadó⁵⁷</td>
<td>New-onset of T2DM</td>
<td>MD + EVOO</td>
<td>4990 80 4271 101</td>
<td>0.01</td>
<td>9</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Salas-Salvadó⁵⁷</td>
<td>New-onset of T2DM</td>
<td>MD + Nuts</td>
<td>4876 92 4271 101</td>
<td>0.126</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Sánchez-Villegas⁵⁶</td>
<td>Depression</td>
<td>MD + EVOO</td>
<td>7715 88 6096 77</td>
<td>0.018</td>
<td>5</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Sánchez-Villegas⁵⁶</td>
<td>Depression</td>
<td>MD + Nuts</td>
<td>6803 59 6096 88</td>
<td>0.003</td>
<td>13</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Singh⁴⁹</td>
<td>Total cardiac events</td>
<td>Indo-MD</td>
<td>499¶¶</td>
<td>39 501 76</td>
<td>&lt;0.001</td>
<td>16</td>
<td>–</td>
</tr>
<tr>
<td>Toledo⁵⁰</td>
<td>First invasive breast cancer</td>
<td>MD + EVOO</td>
<td>7031 8 5829 17</td>
<td>0.027</td>
<td>2</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Toledo⁵⁰</td>
<td>First invasive breast cancer</td>
<td>MD + Nuts</td>
<td>5492 10 5829 17</td>
<td>0.253</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Tuttle⁵¹</td>
<td>Total outcome endpoints***</td>
<td>MD</td>
<td>51 8 50 8</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

*gestational diabetes mellitus (GDM) or preeclampsia.
†Stillbirth, small-for-gestational age (SGA) fetus, or admission to the neonatal care unit.
‡Emergency caesarean section, perineal trauma, pregnancy-induced hypertension and preeclampsia, prematurity, large-for-gestational age, and SGA.
§The number of events was calculated by subtracting the baseline n from the follow-up n count.
¶Cardiac death and non-fatal myocardial infarction (MI).
**In the manuscript, the comparison is reported as ‘significant’ based on the multivariable-adjusted model; however, analysis using Fischer’s exact test does not reveal a significant difference.
††MI, stroke, or death from cardiovascular causes.
‡‡Defined as the achievement of ≥5 risk factor goals: blood pressure <140/90 mm Hg, low-density lipoprotein <2.6 mmol/L, smoking cessation, body mass index <25 kg/m², physical activity of moderate intensity >30 min/day, 3 days/wk (≥6 metabolic equivalents h/wk) and HbA₁c <7% in patients with diabetes.
§§In the manuscript, the comparison is reported as significant (p=0.046); analysis using Fischer’s exact test does not reveal a significant difference.
¶¶In the manuscript, table 4 reports that the size of the intervention group (n) was 999; instead, the count reported in the main manuscript text and the other tables was used for the calculation of the FI (n=499).
***All-cause and cardiac deaths, MI, hospital admissions for heart failure, unstable angina pectoris or stroke.
CV, cardiovascular; EVOO, extra-virgin olive oil; LCD, low-carbohydrate diet; LCMD, low-carbohydrate MD; LFD, low-fat diet; MD, Mediterranean diet; MetS, metabolic syndrome; NAFLD, non-alcoholic fatty liver disease; PAD, periphery artery disease; T2DM, type 2 diabetes mellitus.
discrepancy between two RCTs with high FI indicates that the event rate is not the only parameter influencing the FI. According to Gaudino et al., the FI, *P*-values, events and sample size are mathematically related; however, the type of primary outcome might also have an effect on a trial’s robustness. For instance, the St. Carlos GDM study used two primary outcomes, the first being the incidence of GDM and the second being a composite maternal-fetal score and published the trial’s results in two distinct publications. Although both publications reported significant findings, the first exhibited an FI equal to 4.0 and the latter an FI of 1.0. Composite scores are popular in nutrition research; they are combining distinct outcomes, often resulting in a greater event rate as compared with the use of the ‘component’ outcomes independently. In the present analyses, the use of composite scores did not ensure statistical robustness in all of the trials herein, with many exhibiting low FIs and RFIs (<5).

The Esposito et al. trial also demonstrated a high FI, indicating a robust outcome. However, in this specific RCT the two diets applied by the trialists were not so comparable. In more detail, the intervention arm adopted a low-calorie MD, whereas the comparator group followed a low-fat diet, without any reported restrictions concerning the energy intake. Thus, the observed effects of the intervention arm, and subsequently the high FI, could well have been the result of the prescribed low-calorie diet, as restricted energy intake leading to weight loss has been shown to delay the development of T2DM and subsequently, improve glycaemic control and various coronary factors.

Additionally, it appears that the majority of evidence on MD interventions with dichotomous outcomes is based on the PREDIMED trial, which had a multiarm design. According to Parmar et al., trials concurrently evaluating more than one intervention, like the PREDIMED, have increased chances of finding significant differences even with the use of small sample sizes. Since the FI is based on Fischer’s exact test it can only be applied on 2×2 tables, thus in trials with three parallel arms, distinct comparisons of each intervention with the comparator group were performed for the calculation of the FI/RFI. Exclusion of

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Categorisation of the Fragility and Reverse Fragility Index according to randomised controlled trial design, sample size, compliance assessment method and primary outcomes (n, median and IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Fragility Index</td>
</tr>
<tr>
<td>Masking</td>
<td></td>
</tr>
<tr>
<td>Single blind</td>
<td>21</td>
</tr>
<tr>
<td>Open label</td>
<td>24</td>
</tr>
<tr>
<td>Not reported</td>
<td>2</td>
</tr>
<tr>
<td>Outcome categorisation</td>
<td></td>
</tr>
<tr>
<td>Perinatal outcomes</td>
<td>4</td>
</tr>
<tr>
<td>Outcomes related to the MetS and DM</td>
<td>15</td>
</tr>
<tr>
<td>Cardiovascular outcomes</td>
<td>16</td>
</tr>
<tr>
<td>Outcomes related to NAFLD</td>
<td>3</td>
</tr>
<tr>
<td>Other†</td>
<td>9</td>
</tr>
<tr>
<td>Composite outcome</td>
<td></td>
</tr>
<tr>
<td>Yes‡</td>
<td>15</td>
</tr>
<tr>
<td>No¶</td>
<td>32</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
</tr>
<tr>
<td>&lt;1000 patients</td>
<td>20</td>
</tr>
<tr>
<td>≥100 patients</td>
<td>27</td>
</tr>
<tr>
<td>Dietary compliance assessment methods</td>
<td></td>
</tr>
<tr>
<td>Questionnaires (FFQs, diet scores)/diet history</td>
<td>22</td>
</tr>
<tr>
<td>Diet recalls or records</td>
<td>7</td>
</tr>
<tr>
<td>Biomarkers</td>
<td>18</td>
</tr>
</tbody>
</table>

*Based on the Kruskal-Wallis test. †First incidence of breast cancer, cataract surgery, osteoporotic fractures, return to robustness, depression. ‡Maternal/offspring composite outcome, incidence of metabolic syndrome (MetS), cardiovascular (CV) mortality, composite CV events, optimal CV risk factor control, total cardiac events, total outcome endpoints. §Based on the Mann-Whitney U test. ¶incidence of gestational diabetes mellitus (GDM), atrial fibrillation, heart failure, type 2 diabetes mellitus (T2DM), depression, first invasive breast cancer, steatosis, non-fatal myocardial infarction, cataract surgery, diabetic retinopathy, osteoporotic fractures, need for T2DM medication, reversion to robustness, weight loss, non-alcoholic fatty liver disease (NAFLD) resolution, new symptomatic periphrery artery disease cases, reversion of MetS.

CNC, could not be calculated; DM, diabetes mellitus; FFQ, food frequency questionnaire; IQR, interquartile range.
the trials with three arms however, did not alter the pooled results herein (Median FI: 5, RFI: 7). Accordingly, separation of the PREDIMED comparisons revealed a lack of a significant effect in approximately half of the comparison pairs. When the PREDIMED comparison arms where grouped together and compared against the other trials, the median FIs and RFIs between groups were similar (5 and 7, respectively for both groups), indicating a similar robustness to the rest of MD RCTs. Apart from disputes concerning the randomisation of the PREDIMED sample and the different reported tools used to assess compliance, Correia also noted discrepancies in the medical care offered to the participants, resulting in allocation concealment bias. In parallel, the control group received an intervention of lower intensity for the initial 3 years of the RCT, a corrected problem before completion of the recruitment and analysis of the results. Inevitably, however, a different intervention frequency unmasks participant allocation. Additionally, compliance with the low-fat control diet appeared to be a difficult task in the long run, with the mean fat intake of participants reaching 37.4% of the total energy intake 5 years post intervention. Thus, the control diet did not correspond to a low-fat regime but was rather lower in the fat content compared with the two MD interventions (42% of the total energy intake). Subsequently, more losses to follow-up were recorded in controls, mainly among participants with a worse CVD risk profile at recruitment. This induced further bias towards ameliorated results for the control group, leading to mitigated between-groups differences, and by inference, the bias in the FI. Despite the issues mentioned above, the PREDIMED is an ambitious milestone trial for nutrition research and reanalysis of the data did not reveal differences in the reported results despite the randomisation issues. Given the prolonged intervention duration and the large number of participants, collaborators and outcomes, it is not uncanny that certain aspects of the trial’s design and execution demonstrated issues. Undoubtedly, similar issues might have also been observed in pharmacological trials. On a sidenote, the PREDIMED is probably the only megatrial that has undergone this degree of exhaustive scrutiny, despite the results being unchanged at republication. Moreover, unlike pharmacological trials, the trial aimed in providing evidence to a more traditional and accessible therapy (i.e. diet), without supporting any industry products other than common, ‘healthy’ foods, including olive oil and nuts. According to the authors, these issues should have increased trust to the results. However, for the detailed methodologists, the majority of nutrition research has limitations, whereas for the sceptics, nutrition research is scrutinised for competing against the big Pharma on a pretence of evidence.

For many of the included trials, the calculated low FIs and RFIs were associated with an overall smaller number of events. This problem can be surpassed if greater sample sizes are recruited at baseline, or if we shift the focus towards the execution of pragmatic trials. However, Gaudino noted that it is more ethical to power RCTs in order to produce the required level of evidence using the minimum possible number of participants. Enrolling additional participants might result in stronger evidence against the null hypothesis, however, it might violate the equipoise principle. On the other hand, findings may produce more contradictory results than similar trials, and may also pose further ethical concerns.

An important question arising from the present findings is whether we are receiving the reliable data we are craving for, by performing RCTs, or if we are overlooking important flaws of either the nutrition science, or the methodology applied in trials examining the MD. However, the present study did not aim in examining the importance or the effectiveness of the MD as a therapeutic dietary regimen. The low robustness calculated herein indicates that even the best level of primary MD evidence proving causality, namely the RCTs, can fail to reach the standards one would expect. Recently, a study assessing the FI of clinical nutrition trials revealed a low FI. According to Zeilstra, many nutritional RCTs yield ambiguous results, which is why the RCT design is often considered ‘ill-suited’ for nutritional research. Additionally, that most trials are based on different analyses of the same landmark protocol (PREDIMED), bias and limitations of the trial are inevitably reproduced in every publication. Subsequently, any synthesis of related RCTs, although it may present low heterogeneity, carries an inherited risk of extrapolated findings. To nutrition’s defence however, lower median FI compared with that of MD interventions has been reported in perioperative, anaesthesiology, plastic surgery, and critical-care medicine RCTs, as well as among paediatric orthopedic and appendicitis trials. Nevertheless, the synthesis of these trials for recommendations formulation consists of a common practice in the fields mentioned above, as in the science of nutrition.

On the flip side, RCTs with MD interventions and continuous primary outcomes demonstrate significant findings while supporting the health benefits of adhering to the MD prototype. However, similarly to the Esposito trial, control interventions are not always comparable, with a tendency to favour the MD arms. This is why, to verify the health effects of MD adherence and advocate for its prescription, superiority trials with continuous primary outcomes should be performed, comparing the MD to other healthy diet regimens instead of the usual diet of participants or dietary advice only.

Although the current results indicate that as far as trials with dichotomous outcomes are concerned, the evidence on the MD entails some limitations, several other factors must also be considered before treating the MD with contempt. For instance, assessment of the participants’ adherence to the dietary intervention, often relies on short dietary indexes instead of more objective measures, and consists of an important component of a nutrition RCT. Moreover, the Hawthorn effect (individuals modify an aspect of their behaviour in response...
to their awareness of being observed) is apparent in all of nutrition research; thus, compliance and assessment are not always accurate. RCTs are often used to guide clinical practice and are sometimes incorporated in clinical practice guidelines intact or after synthesis, using systematic reviews and meta-analyses. Given the demand for evidence-based nutrition recommendations, the results suggest that the formulation of recommendations promoting the MD based on RCTs should be performed with caution. Thorough examination of the American College of Gastroenterology guidelines revealed that most RCTs used to guide recommendations regarding Crohn’s disease relied on a small number of superior events for ‘securing’ statistical significance. Often, the FI coincided with the drop-outs reported in some trials. This is why, reporting the FI has also been suggested for systematic reviews and meta-analyses, to understand the fragility of the presented associations and identify possible misuse of the P-value. The present study aimed to pinpoint another issue requiring the attention of scientists when performing nutrition trials, namely the FI. Meticulous care in the trial design, sample size and execution can improve the FI of nutrition trials and aid in upgrading the science of nutrition, as succinctly pointed out by other researchers.

Another important issue in nutrition research is that often, detailed definitions of the interventions are not reported. This is also the case with the MD. Although the label MD is a generic term used to describe the diet of inhabitants around the Mediterranean basin, according to Martínez-González, what constitutes the MD and its key determinants differs even among ‘experts’ worldwide. Martínez-González noted that the discrepancies in the MD definition consist of a major problem, especially for intervention studies. As a result, except for the RCTs included herein which were stemming from the same protocol, like the PREDIMED, the remaining trials have most probably used different definitions of the MD. For instance, Singh and associates used a National Cholesterol Education Program modification of the MD, whereas Greenberg et al reported following Professor Willet’s definition of the MD. This indicates that differences may exist even under the same intervention label, and these may well induce inconsistencies and bias in the reported outcomes.

Undoubtedly, one important limitation of the study stems from the relatively small number of RCTs with a dichotomous primary outcome included in the analyses. However, one should consider that the total number of RCTs examining MD interventions is rather small; additionally, in the present study, RCTs were selected based on a systematic search strategy; thus, the results reflect the actual number of available MD-RCTs fulfilling the study’s criteria and being indexed in the PubMed database. An additional limitation is that the publication of many RCTs predated the CONSORT guidelines; thus, few important characteristics have not been reported. In parallel, in the case of MD RCTs, as in the majority of nutritional epidemiology, diet adherence and intake rely on not so precise exposure assessments—mainly self-reported information—with an increased potential for confounding.

Moreover, due to the small number of retrieved trials, it was not possible to correlate the FI with individual study characteristics, or to perform additional statistical analyses. As already mentioned, the use of broad research topics for the assessment of the FI/RFI, as seen herein with the MD, is common in the literature. Although such studies result in pooling a greater number of RCTs, they also tend to mix many studies with non-comparable aspects, including participant age, health status, study question, outcomes categories, etc. In an effort to correct the heterogeneity observed in the included trials, we also calculated the FI and the RFI after allocating the RCTs based on sample sizes, masking, or outcomes categories. However, these analyses failed to reveal differences, with the only observed significant finding involving the different RFI among RCTs using different methods to assess dietary adherence. Therefore, in the pooled sample of RCTs included herein, differences in sample size, outcomes categories or masking had a minimal effect on the FI and the RFI. Nevertheless, a larger pool of RCTs might have produced different results.

Limitations of the FI include the fact that its calculation is based on the Fischer’s exact test, which is considered as stricter and more prone to type II errors when compared with the χ² test. Additionally, as already mentioned, it can only be applied to dichotomous outcomes, whereas the majority of nutrition research tends to examine continuous outcomes. Furthermore, the lack of standardised cut-offs for categorising RCTs as either robust, or fragile, is evident. According to Andrade, the most important limitation of the index concerns the use of the much decried statistical threshold (p<0.05) for determining the significance of a study’s outcome. However, one should consider that the FI uses the same threshold applied in the published RCTs and that additionally, the FI is highly correlated to the P-value of a trial, with a significance closer to 0.05 indicating a lower FI. Moreover, although Walsh suggested calculating the index in time-to-event data—as performed in the current analysis—several researchers raised concerns, claiming that it cannot account for the effect of time. Nevertheless, as Charilau promptly noted, the FI can offer a measure of the validity of an RCT, especially in trials where the number of participants lost to follow-up, exceeded the FI of the trial. More recently, in a collective effort to optimise patient care, the routine use of the FI has been recommended for the development of all clinical practice guidelines, with incorporation of the results in the GRADE (Grading of Recommendations Assessment, Development and Evaluation) format.
CONCLUSIONS

In summary, the present study reveals that, when adhering to good scientific principles, one discerns that even in the top tiers of evidence hierarchy, research on the MD may lack robustness, setting concerns for the formulation of nutrition recommendations in a wider context. A collective effort is required to promote the science of nutrition in an evidence-based manner. Despite the mediocre robustness of RCTs with MD interventions, the findings herein do not overlay on the importance of the MD on health or as a UNESCO accredited intangible cultural heritage. Nevertheless, it appears that our quest for an ideal diet for all could prove horses for courses, and a more personalised approach may be required for both health attainment and ameliorated disease outcomes. As Correia noted ‘enthusiasm regarding the MD may not be proportional to the level of evidence’ and this might lead to allegiance bias and an imbalance between expectancies and evidence.

Author affiliations

1 Unit of Reproductive Endocrinology, 1st Department of Obstetrics and Gynecology, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Central Macedonia, Greece
2 Rheumatology and Clinical Immunology, Faculty of Health Sciences, University of Thessaly, Larissa, Greece
3 Nutritional Sciences & Dietetics, Faculty of Health Sciences, International Hellenic University, Thessaloniki, Greece
4 Medical School, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Central Macedonia, Greece
5 Department of Endocrinology, Diabetes and Metabolism, Hippokration General Hospital of Thessaloniki, Thessaloniki, Central Macedonia, Greece
6 Laboratory of Histology and Embryology, Medical School, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Central Macedonia, Greece
7 Division of Transplantation, Immunology and Mucosal Biology, MRC Centre for Transplantation, School of Medical Education, King’s College London, London, UK
8 Meletios P Nigdelis @melnigdelis, Xenophon Theodoridis @x_theodoridis and Dimitrios P Bogdanos @dbogdanos

Acknowledgements The present work has been submitted for presentation at the 10th Scientific Congress of the Medical School, Aristotle University of Thessaloniki (2021).

Contributors MGG and DGG developed the research question and finalised the paper inclusion. MPN, MGG, KG and XT performed the search and extracted the data. AT and TP participated in the search. KG designed the protocol. XT, MGG and DGG interpreted results, drafted and finalised the manuscript. MPN included trials, with some help from MGG and supervision from DGG and DPB. MPN and MGG performed the statistical analyses, KG, MPN, XT, TP and MGG prepared tables and figures. DGG, DPB and MGG interpreted results, drafted and finalised the manuscript. All authors read the manuscript, contributed comments to its revision, and have approved and agreed to the final version. MGG submitted the manuscript, and DGG is responsible for the overall content as guarantor.

Funding The present research was funded by the “MSc in Health and Environmental factors”, Medical School, Aristotle University of Thessaloniki.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Although most of the data are already presented in the manuscript text, they are also available upon reasonable request. For expression of interest, please contact Prof. Dimitrios G. Goulis.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs

Maria G Grammatikopoulou http://orcid.org/0000-0003-4167-6595 Meletios P Nigdelis http://orcid.org/0000-0003-3965-6833 Xenophon Theodoridis http://orcid.org/0000-0001-9810-7583 Konstantinos Gkiouras http://orcid.org/0000-0002-6407-8494 Antigoni Tranidou http://orcid.org/0000-0001-6406-8424 Theodora Papamitsou http://orcid.org/0000-0002-0709-5407 Dimitrios P Bogdanos http://orcid.org/0000-0002-9697-7902 Dimitrios G Goulis http://orcid.org/0000-0002-5005-1995

REFERENCES


21 Ioannidis JP. We need more randomized trials in nutrition-preferably large, long-term, and with negative results. Am J Clin Nutr 2016;103:3385–6.


23 It’s time to talk about ditching statistical significance. Nature 2019;567:283.


