Evaluation of Mediterranean diet adherence scores: a systematic review

A Zaragoza-Martí, 1 MJ Cabañero-Martínez, 1 JA Hurtado-Sánchez, 1 A Laguna-Pérez, 1 R Ferrer-Cascales 2

ABSTRACT

Objective The aim of this review was to evaluate the conceptual suitability, applicability and psychometric properties of scores used internationally to measure adherence to the Mediterranean diet (MD).

Design This was a systematic review to identify original articles that examined some aspects of the conceptual suitability, applicability or psychometric properties of the MD adherence score. Electronic searches were carried out on the international databases MEDLINE, Scopus, Web of Science and EMBASE (from January 1980 to 31 December 2015).

Eligibility criteria for selecting studies The study included original articles that examined some aspects of the conceptual suitability, applicability or psychometric properties of the MD adherence score. The studies where MD adherence scores were administered but did not bring forward any evidence about their performance related to conceptual suitability, applicability or psychometric properties were excluded.

Data extraction Information relating to the scales was extracted in accordance with the quality criteria defined by the Scientific Advisory Committee of the Medical Outcomes Trust for measurement of health results and the quality criteria recommended by Terwee: (1) conceptual, (2) applicability and (3) psychometric properties. Three authors independently extracted information from eligible studies.

Results Twenty-seven studies were identified as meeting the inclusion criteria, yielding 28 MD adherence scores. The results showed that evidence is scarce and that very few scores fulfilled the applicability parameters and psychometric quality.

Conclusions Scores measuring adherence to MD are useful tools for identifying the dietary patterns of a given population. However, further information is required regarding existing scores. In addition, new instruments with greater conceptual and methodological rigor should be developed and evaluated for their psychometric properties.

INTRODUCTION

Several epidemiological studies have evaluated the relationship between health and food intake. 1–6 Specifically, various population surveys and clinical trials provide evidence that diets that are high in fruits, vegetables, legumes, whole grains and fish, and moderate in dairy intake, are associated with lower incidence of chronic diseases. 4–7–10

The Mediterranean diet (MD) is characterised by a high intake of plant-based foods (vegetables, legumes, fruits, nuts, cereals (mainly whole grain)), olive oil as the main source of fat, moderate amounts of dairy (yoghurt and cheese), low or moderate consumption of fish and meat, moderate consumption of wine consumed with meals, and an active lifestyle. 11–14 Although the various geographical regions of the Mediterranean have different diets, influenced by sociocultural, religious or economic factors, among others, it can be assumed that these diets are variations of the same MD diet. 15–16

Various longitudinal studies have analysed the benefits of MD in comparison with other types of diet. 17–23 These studies have shown that people with good adherence to MD have...
a better quality of life and greater life expectancy, along with a decreased prevalence of chronic diseases such as certain types of cancer, type 2 diabetes, and cardiovascular or neurodegenerative diseases. Specifically, the protective role of MD has been attributed to the high intake of plant-based foods, along with a moderate consumption of wine, fish and dairy, and a high intake of monounsaturated fatty acids in lieu of saturated and trans fatty acids, which are linked with an elevated antioxidant capacity. Therefore, it is important to ascertain the degree of adherence to MD through accurate measurement tools such as dietary scores based on the frequency of pattern-consistent and pattern-inconsistent food consumption, as well as compliance with recommended intake.

Evidence shows that dietary scores are useful tools to evaluate the degree of adherence to MD and its benefits in regard to health. Scores are composite constructs based on dietary components, combining foods and nutrients to obtain valid operational variables that analyse the association between the quality of diet and its health effects. Several scores are used to measure the degree of agreement with MD. The first and most widely used score was created by Trichopoulou et al in 1995. This score evaluates concordance with the dietary pattern, by assigning one point when the intake of protective foods is higher than the median, in the study/sample population, or when the consumption of non-protective foods is lower than the median, and zero in the opposite situations. Other scores based on MD have been created for use in different geographical populations, for populations with different underlying physiological states, so that alternate foods can be incorporated into and/or accounted for within the canonical pattern.

The characteristics of MD scores have been reviewed in different studies. However, the quality of these instruments, which is fundamental to ensuring their valid and reliable application, has not been analysed. The heterogeneity of MD adherence scores raises the potential for disparity in analyses, as well as confusion as to which specific score to choose. Therefore, to be able to select a good instrument, one must first know the quality criteria it offers. Knowledge of such criteria is imperative for the accurate use of the instrument. According to the Scientific Advisory Committee (SAC) of the Medical Outcomes Trust, eight quality criteria must be established, corresponding to three groups of information: conceptual suitability (conceptual and measurement model, cultural and linguistic adaptation); applicability (demands of the administrator and respondent, alternative forms, interpretability); and psychometric properties (reliability, validity and responsiveness).

For this reason, the aim of this review was to evaluate the conceptual suitability, applicability and psychometric properties of MD adherence scores used internationally.

### METHODOLOGY

#### Search strategy

To obtain original documents, electronic searches were carried out using the following international databases: MEDLINE, Scopus, Web of Science and EMBASE. The search strategy was designed to obtain original studies about the development or validation of scores measuring adherence to MD, published until 31 December 2015 (January 1980 to 31 December 2015). This strategy focused on combining the following keywords: Mediterranean diet, score and adherence, and terms associated with the psychometric properties of instruments (validity, quality and reproducibility). In order to increase the sensibility of the search strategy, searches were conducted using the thesaurus of each of the databases selected and keywords—in the title and abstract—associated with the search terms (figure 1). The electronic searches were complemented by manual searches in international journals with regard to their relevance and frequency in the publication, by new searches on PubMed under the names of the identified MD score and under the names of the authors who had created or adapted them, and by the references of the articles that complied with the inclusion criteria. Abstracts from congresses and grey literature were excluded.

<table>
<thead>
<tr>
<th>Search Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (mediterranean diet[Title/Abstract]) OR mediterranean diet[MeSH Terms]</td>
</tr>
<tr>
<td>2. (adherence [Title/Abstract])</td>
</tr>
<tr>
<td>3. (score [Title/Abstract]) OR (index [Title/Abstract])</td>
</tr>
<tr>
<td>4. ((quality) [Title/Abstract]) OR (validity[Title/Abstract]) OR reproducibility of results [MeSH Terms] OR reproducibility of results [Title/Abstract] OR psychometrics [MeSH Terms] OR psychometrics [Title/Abstract]</td>
</tr>
<tr>
<td>5. 1 AND 2 AND 3 AND 4</td>
</tr>
</tbody>
</table>

Figure 1 Search strategy using MEDLINE for studies on the evaluation of Mediterranean diet adherence scores. Search was conducted on MEDLINE with the appropriate search terms used for the other databases.
Inclusion criteria
All original articles that examined some aspects of the conceptual suitability (conceptual and measurement model, cultural and linguistic adaptation), applicability (demands of the administrator/respondent, alternative forms and interpretability) or psychometric properties (reliability, validity and responsiveness) of the MD adherence score in English or Spanish published until 31 December 2015 (from January 1980 to 31 December 2015) were included.

Exclusion criteria
The studies where MD adherence scores were administered but did not bring forward any evidence about their performance related to conceptual suitability, applicability or psychometric properties were excluded.

Selection of studies
Two reviewers (RF-C and AZ-M) assessed the titles and abstracts to determine their inclusion or exclusion from the review. The reviewers worked independently, and if they were in disagreement a third reviewer (MJ-C-M) would resolve the disagreement or recommend reading the whole article.

Data extraction
Information was extracted by the same researchers (MJ-C-M, RF-C and AZ-M), who had independently carried out the selection of original articles, resolving disagreements through consensus with a third person. The information extracted was divided into two sections: information about the characteristics of the study and the sample, and information about the measurement scales. The first section included the characteristics of the study and the sample (inclusion criteria, sample size and origin of the population).

Information relating to the scales was extracted in accordance with the quality criteria defined by the SAC of the Medical Outcomes Trust for measurement of health results and the quality criteria recommended by Terwee.36–39 In order to facilitate understanding, the health results and the quality criteria recommended by the SAC of the population).

and training of interviewers; +++: normative data); (3) reliability (?: not reported or weak associations of some aspects of internal consistency reported; +: alpha coefficient of internal consistency, or intrarater or inter-rater reliability reported; ++: alpha coefficient or interclass correlation coefficients (ICC) or correlated coefficient >0.70; and (4) validity (?: not reported; +: evidence from criterion or construct validity; ++: evidence from criterion and construct validity).

RESULTS
Search results
A total of 56 articles met the inclusion criteria, which were reduced to 52 once the duplicates had been removed (figure 2). In addition, 19 of these articles were excluded after reviewing the title and the abstract because they did not meet the inclusion criteria. Finally a further six articles were excluded because they did not use specific MD adherence scores in their methodology. Therefore, 27 articles were included in the review, from which 28 MD adherence scores were used.

Characteristics of included studies
The designs of the studies included were principally observational (12 cohort studies,14 16 26 28–31 43–47 1 case and control study,34 14 descriptive studies6 11 12 29 32 33 48–55 and 1 intervention study36). A total of 17 studies focused on the general population,6 14 26 29 32–34 46–50 52–56 3 on the elderly,15 43 45 2 on children,11 12 1 on university students16 and 1 on pregnant women.31 Finally, three of them did not indicate the target population of the scores.16 28 44 With respect to sample size, the scores created by Trichopoulou et al14 43 were developed using large samples: 22 043 and 74 607 people, respectively. There were three studies with a sample size of < 150 people.29 51 56

Conceptual suitability
Online supplementary table 2,3 summarise key data regarding the conceptual suitability of the different scores: the context in which they were applied, content validity and cross-cultural adaptation process. The scores were listed according to their conceptual model and measurement. The majority of the scores (n=18)6 11 14 16 26 29–34 43–45 48 49 51 were based on positive and negative components of MD. Five of them were based on the structure of the MD food pyramid,28 32–34 52–56 three on the general characteristics of MD40 47 55 and one on the Diet Quality Index.12 As a fundamental model, the scores created by Trichopoulou et al14 30 43 have been the most widely used, with six scores being created on the basis of their components.16 26 29 31 45 50

Although there is no consensus on the meaning of the ratings, as a general rule, interpretation of these scales is positive for healthy items and negative for unhealthy items, with high scores indicating good adherence to the MD and low scores, poor adherence. Only the scores created by Scali et al80 and Gerber40 provide inverted
scores, where high scores indicate low adherence and low scores indicate good adherence (online supplementary table 2).

The majority of the scores were developed in Mediterranean countries: Spain (n=14), Spain (n=3),6 14 26 29 51–54 Greece (n=3),11 12 16 26 29 31–34 47 50 53 54 Italy (n=2)6 46 47 and France (n=2).48 49 The remainder were developed in Canada (n=1),56 other European countries (n=3),43–45 Japan52 55 and the USA (n=2)28 55 (see online supplementary table 2).

Regarding the context of application (online supplementary table 3), 12 of the 28 scores analysed were applied to the general population,16 26 45–47 46–54 56 6 in primary care,6 26 32 43 48 55 3 in hospital care,31 33 34 6 in the community,6 11 14 28 30 55 and 1 in sports clubs.12 The scores developed by Panagiotakos et al6 and Woo et al55 are used in the context of primary care and also in the community.

None of the MD adherence scores detail the process of cross-cultural adaptation. The majority of the scores come from the one Food Frequency Questionnaire (FFQ) previously validated for the population studied; however, in the original studies of this instrument (FFQ), the process of cross-cultural adaptation has not been detailed.

With regard to content validity, the majority of scores based on negative and positive components6 14 26 29 31 43 45 50 are created in function of the scores developed by Trichopoulou and colleagues.30 Scores of the MD pyramid are based on the pyramid elaborated by Bach-Faig and colleagues.57 The rest of the scores are founded in general references of the MD pattern.

**Applicability**

Relating to the applicability of the MD adherence scores, with the exception of the score created by Woo et al,55 who did not specify the method of administration, all diet questionnaires were administered by trained interviewers. Regarding the source of information, all of the scores were answered by the patients/participants (not by a
of the scores provided an estimation of a statistic capable of measuring effect size. Only the scores developed by Goulet et al.6,14 26 48 55 examined the effect of a nutritional intervention, in which MD adherence scores increased significantly from 21.1±3.6 in week 0 to 28.6±4.4 (P<0.001) after 6 weeks of intervention.

Online supplementary table 7 presents the MD summary scores. Only four scores did not provide any information about the cross-transcultural process.14 31 32 47 The scores developed by Panagiotakos et al.6,14 Scali et al.48 Gerber50 and Woo et al.55 obtained the best evaluations in terms of applicability. The score created by Sotos-Prieto et al.54 was the instrument with the most and best evidence on reliability. Information about validity was provided for most of the scores, but concurrent and predictive validity were only reported for the scores created by Panagiotakos et al.6 Buckland et al.65 Sotos-Prieto et al.54. However, only the study by Sotos-Prieto et al.54 provided information about reliability.

**Discussion**

The review conducted here included 27 references and identified 28 MD adherence scores used internationally. The evidence obtained from these studies has been evaluated based on conceptual suitability, applicability and psychometric properties. The results reveal that evidence is scarce, and that very few scores fulfill psychometric properties and applicability parameters typically associated with scales/indices. The scores developed by Panagiotakos et al.,6 Buckland et al.65 and Sotos-Prieto et al.54 provide the most information. However, as with the other scores analysed, none of them provide complete information about the process of transcultural adaptation used. The scores reviewed here only specify that a previously validated FFQ for the original population has been used, but do not provide the transcultural adaptation of these dietary questionnaires (translation, back translation and pilot study). The Scientific Committee of the Medical Outcomes Trust39 considers cultural and linguistic adaptation to be an especially important criterion in achieving linguistic and cultural equivalence with an original instrument.

Applicability is one of the sections that present the most information gaps. None of the scores report on normative data, and only five of them14 25 48 55 provide detailed information about the administration process using photographic and visual material to obtain information as close to reality as possible.

The data about reliability are the most deficient. To ascertain the degree to which all the items on a scale measure the same construct, internal consistency must be measured. In this case, the score created by Sotos-Prieto et al.54 is the only one that provides information about this topic, through the Cronbach alpha value. The degree of association between the scores obtained and the items included on the instrument has been taken into account,
but this information cannot be considered a quality item-
test measure of reliability. Regarding reliability data, only
the two scores created by Benítez-Arciniega et al20 display
test–retest reliability and equivalence reliability.

Validity was the most widely reported property. Only
the scores created by Benítez-Arciniega et al20 did not
include any information about validity. In the scientific
literature, there are different gold standards to evaluate
criterion validity, such as clinical and biological markers
for concurrent validity, and adverse events for predictive
validity. However, the best gold standard, ‘observation of
food intake’, has not been used in any of the studies. In
some of the studies analysed,26 31 the gold standard used is
the score created by Trichopoulou et al,30 obtaining agree-
ment levels of close to 70% with the original, considered
here to provide construct validity. This one was the first
score used to measure levels of adherence to MD, but it
cannot be considered a gold standard, since there is new
evidence indicating changes in food and diet patterns.
It should also be pointed out that no confirmatory anal-
ysis was conducted in relation to the structure of the
instruments.

It has been consistently demonstrated that MD helps to
protect against cardiovascular disease, inflammatory and
metabolic diseases, as well as numerous chronic degener-
avative diseases1 2 35 58–63; nevertheless, the protective effect of
MD is very different across the studies.35 64 Consequently,
a large number of MD adherence scores are being created
to ascertain the relationship between diet and health.
However, recent publications indicate that some of these
scores do not offer strong predictive capacity regarding
mortality or disease, thus questioning the quality.13 64 65
This observation is borne out by the findings of this study,
which has shown that the majority of the scores analysed
are lacking in information about the quality attributes of
the scales.

For all of the above reasons, greater attention must be
paid to the way in which these scores have been created.
First, a common criterion should be established to iden-
tify the components that make up MD. Second, different
elements need to be unified: the number of components
(nutrients, foods or food groups), classification catego-
ries for each population, measurement scale, statistical
parameters (mean, median, tertiles and so on) and the
contribution of each component (positive or negative) to
the score total.35 53 60 67 Finally, given the great heteroge-
nenity of MD in different countries, further confirmatory
analyses are required using biomarkers with a view to vali-
dating said dietary pattern.

Strengths and limitations
Although the data are conclusive regarding the lack of
quality of MD adherence scores and the need to improve
the measurement of MD adherence, it is important to take
into consideration the limitations of this review, which are
related to the process of bibliographic searches, derived
from the electronic search and retrieval of documents.
In order to control this limitation, multiple synonyms
of the search terms were used, and complementary
searches of prestigious journals and bibliographic refer-
ences were also conducted. Furthermore, this review
only took account of studies wherein the main objective
was to develop or examine data about the applicability
or psychometric properties of an MD adherence score.
It could produce an underestimation of the predictive
and/or concurrent validity, which are the most frequently
analysed in longitudinal studies on MD adherence scores.

In conclusion, the use of scores to measure adherence
to MD is a very useful tool for identifying the dietary
patterns of the population. However, our results point
tout that fewer of the analysed scores suit the quality
criteria. The scores developed by Panagiotakos et al,6
Buckland et al26 and Sotos-Prieto et al31 have obtained better
evidence, although they have not been considered
as gold standard because they do not fit all of the quality
criteria. As a consequence, it could be possible that the
employed scores to evaluate the relationship between
MD and health do not present a good predictive ability,
originating significant bias in the obtained results. For
all these reasons, further information is required about
the scores that currently exist, and/or new instruments
with better concept grounded must be developed. Future
research should focus on improving the psychometric
properties of the MD adherence scores and analysing the
concordance between these instruments in compliance
with the normative quality criteria.

Contributors Conceived and designed the experiments: AZ-M, MJC-M, RF-C.
Analysed the data: AZ-M, MJC-M, RF-C, JAH-S, AL-P. Wrote the paper: AZ-M,
MJC-M, RF-C, JAH-S, AL-P. Data interpretation and critical revision of manuscript:
AZ-M, MJC-M, RF-C, JAH-S, AL-P. All authors reviewed and approved the
manuscript.

Funding This research received no specific grant from any funding agency in the
public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

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 and the use is non-commercial. See: http://creativecommons.org/
licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the
article) 2018. All rights reserved. No commercial use is permitted unless otherwise
expressly granted.

REFERENCES
1. Sofi F, Cesari F, Abbate R et al. Adherence to Mediterranean diet and
2013;39:335–42.
Mediterranean diet reduces the risk of metabolic syndrome: a 6-year
type 2 diabetes: cardiometabolic benefits. Endocrine 2017;56.


38. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.


