Current Medical Research and Opinion

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/icmo20

Delphi, non-RAND modified Delphi, RAND/UCLA appropriateness method and a novel group awareness and consensus methodology for consensus measurement: a systematic literature review

Ravi Jandhyala

To cite this article: Ravi Jandhyala (2020): Delphi, non-RAND modified Delphi, RAND/UCLA appropriateness method and a novel group awareness and consensus methodology for consensus measurement: a systematic literature review, Current Medical Research and Opinion, DOI: 10.1080/03007995.2020.1816946

To link to this article: https://doi.org/10.1080/03007995.2020.1816946

CURRENT MEDICAL RESEARCH AND OPINION
https://doi.org/10.1080/03007995.2020.1816946
Article ST-0595/1816946

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Delphi, non-RAND modified Delphi, RAND/UCLA appropriateness method and a novel group awareness and consensus methodology for consensus measurement: a systematic literature review

Ravi Jandhyala

Medialis Ltd, Banbury, UK

ABSTRACT

Introduction: Increasing demand for reliable evidence in patient care and its delivery has necessitated the development of several approaches for generating quality evidence. In particular, the solicitation of expert opinion has been recognised as a reliable data collection method. However, there are variations and limitations in study approaches using expert opinion as a method of data collection, thereby necessitating the development of a standardised, novel consensus method.

Methods: A systematic literature review was conducted to assess the characteristics of all studies utilising a "Delphi" or "Modified Delphi" methods between January 2008 and December 2018. A search framework was developed, and the review was conducted in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Results: In total, 764 studies met the study inclusion and exclusion criteria and were included in the review. Heterogeneity on core defining characteristics of the constituent study types was observed in this control group. These 764 studies were compared against the four case studies using the Jandhyala method. Four key themes were identified and discussed: Assessment of Forced or Observed consensus, Assessment and reporting of item Awareness and advisor awareness, Minimum expert engagement profile, and Efficiency of Minimum Engagement Profile.

Conclusions: Existing consensus methodologies have undergone significant modifications by successive authors over time, including ones contradicting core principles where an original method had been defined. The Jandhyala method for generating group consensus and awareness is unique in observing consensus and measuring awareness of subject matter across experts. The Jandhyala method also improves upon the traditional Delphi-style methodologies, through the introduction of new insights into awareness of subject matter in the expert group. A wider application of the Jandhyala method is required to corroborate findings from this research.

ARTICLE HISTORY
Received 14 May 2020
Accepted 24 August 2020

KEYWORDS
Delphi; Jandhyala method; novel consensus method; case studies; expert opinion

Introduction

In medical and healthcare research settings, there is a strong and continuous need for high-quality evidence around all aspects of medicines and diseases treatment. Information on areas such as disease epidemiology, disease progression, drug safety and optimal dosing regimens, are important to inform stakeholders (such as Healthcare Professionals (HCPs), regulators and payors), and allow patients to receive the optimal level of care[1–4]. In view of this, results from meta-analyses of Randomised Clinical Trials (RCTs) are considered the highest level of evidence. While RCTs are generally accepted as the "Gold Standard" and are used by regulators to inform decisions on approval of marketing authorisation of new medicinal products, payors for reimbursement of those same medicines are less likely to rely on evidence from RCTs[5].

RCT's are very resource intensive, time consuming, and often of limited scope. In many cases, RCTs will have been designed solely to meet the needs of the regulator to obtain a marketing authorisation approval, with little consideration given to meeting the evidence needs of the remaining downstream stakeholders, such as the prescriber or payor[6]. Therefore, RCTs are not typically a viable option to answer the full range of "downstream" research questions that will exist around any new medicine. This raises a case for the concurrent utilisation of other, complementary methods of evidence generation, potentially from sources occupying lower levels of the evidence pyramid. In this setting, the solicitation of expert opinion has been recognised as an appropriate method of generating reliable evidence and, while it is of a lower level than clinical studies, it is suited to the rapid filling of key data gaps to meet immediate needs.

CONTACT Ravi Jandhyala ravi@medialis.co.uk Medialis Ltd, 13 Horse Fair, Banbury OX16 0AH, UK

Supplemental data for this article is available online at https://doi.org/10.1080/03007995.2020.1816946.

Currently, expert opinion is developed *via* consensus generating methodologies, such as Delphi or its derivatives, the RAND/UCLA appropriateness method (RAM), or non-RAND modified Delphi (nRMD) methods[7,8]. However, it has been suggested that these methodologies are subject to certain limitations. These limitations include high attrition of experts (due to potentially infinite consensus rounds), the notion of "forcing" of consensus (*via* process fatigue and/or "railroading" of opinion by dominant personalities), and a lack of understanding of the levels of awareness that the recruited experts have on the subject being addressed[9]. Furthermore, the evolution of the Delphi approach since its original development in 1958 has led to numerous modifications, with some methodologies differing fundamentally from the original[10,11]. This evolution of Delphi and its derivatives over time has gone largely unchallenged, with little formal reporting of any changes made. Therefore, a true understanding of the degree of variation in methodologies calling themselves "Delphi" (or a derivative) is needed to define these approaches and enable comparisons to be made between them. This will aid in the definition of a standard against which a new method can be compared.

The Jandhyala Method[12], which has been recently published, is a novel methodology for observing group awareness and consensus which potentially overcomes many of the above limitations with the Delphi methodology and its derivatives[12]. As a new option, a formal appraisal against the current standard is needed. This formal appraisal will be expected to critically appraise the characteristics of the new method, taking into consideration areas where it may represent a differentiated alternative or an improvement to the standard. If successfully differentiated and deemed to be a more attractive option, there may be sufficient justification for its wider adoption.

This research has dual objectives. Firstly, it aims to appraise the existing consensus generating methodologies (Delphi, RAM and nRMD) in the healthcare setting for consistency, allowing the definition of a "standard", or a control group, for further analysis. Secondly, it will conduct comparisons between the new Jandhyala method and the defined "standard". These two objectives will be met using a systematic review, focussing on:

1. the uniqueness of the study objective in observing consensus,
2. measurement of awareness of experts recruited,
3. profile of engagement with experts, and
4. overall efficiency in terms of the number of rounds, duration and attrition of experts.

Methods

A systematic literature review was conducted to assess the characteristics of all studies utilising a "Delphi" or "Modified Delphi" (including RAND-UCLA appropriateness methods) between January 2008 and December 2018 following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Study search

A full systematic literature search was carried out electronically in MEDLINE *via* PubMed and on Google Scholar on 19 December 2018 using combinations and synonyms of the keywords: "Delphi" and "Healthcare". Search strings were built using the PICO framework so that studies could be scanned for relevance. While common practice is to search the first 50–100 hits (5–10 pages) in search engines, a review finds that the majority of grey literature appears at page 20–30 of the results in Google Scholar, with the highest volume of grey literature starting at page 35. Consequently, the first 350 hits were examined for the appropriate data[13].

Study inclusion and exclusion criteria

Table 1 lists the full inclusion and exclusion criteria applied to the studies identified in the literature review. Identified studies underwent a two-round screening process, firstly an initial screen of the title and abstracts, followed by an in-depth full-text review for studies passing the title and abstract screen.

Data extraction

All studies selected for inclusion were reviewed separately by eight researchers from Medialis Ltd, and all relevant data were extracted. The total list of included papers was divided into four groups, and data in the papers in each group were independently extracted and quality checked by two researchers. A list of all the studies used during the literature review is presented in Supplement 1. Any queries which could not be resolved between the two researchers were adjudicated on by a third researcher. A list of all items extracted during the literature review is presented in Appendix 1.

A bespoke quality and validity assessment checklist was developed by the author to both accommodate the non-clinical nature of the literature included in the review and address the absence of an independent checklist dedicated to appraising expert opinion specifically. These results are presented in Appendix 1 as "Quality Reporting".

Data analysis

Heterogeneity of existing consensus generating methodologies (control group)

The main defining characteristics of the three Delphi-based methods are listed in Table 2. They differ in several key aspects, namely: whether expert anonymity is preserved, the degree to which feedback on the responses to the previous consensus round is provided, whether systematic literature review was conducted, and whether the expert was allowed to influence or modify the list on which consensus is being solicited between rounds[8,10]. Delphi and RAM, a modification of Delphi, have referenced descriptions of their method[8,10]; however, this is not the case with nRMD. Delphi uses serial rounds of consensus assessment *via* questionnaires, between

Table 1. Showing the inclusion and exclusion criteria applied to all studies identified during the systematic literature review (Page 5).

Inclusion criteria	Exclusion criteria
Original, published research articles	Non-original research articles
Studies involving solicitation of expert opinion (with or without literature reviews) using Delphi or modified Delphi approaches	Non-healthcare related fields (such as economics, management, politics, etc.)
Healthcare field assessing disease, disability or health outcomes with or without an association to any intervention	Studies not describing Delphi or modified Delphi methodologies
Studies detailing the characteristics of the Delphi or modified Delphi approach	No full text available
	Letters
	Editorials
	Protocols
	Systematic reviews
	Narrative reviews
	Textbooks
	Non-English language publications

Table 2. Showing definitions for Delphi-based methods (Page 7).

Aspects characterising methodology	Delphi based methods		
	Delphi	nRMD	RAM
Anonymity (No F2F meeting)	Yes	No	No
Feedback	Yes (controlled)	NA	Yes (Own + summary from all)
Systematic literature review	No	Yes	Yes
Expert allowed to modify the list	NA	NA	Yes

which controlled feedback is provided on responses which enable experts to alter their responses in subsequent rounds. Importantly, there is no face to face (F2F) meetings between experts, preserving anonymity, and no literature review was described.

RAM was described by the RAND Institute and UCLA for a well-circumscribed and specific purpose, in contrast to the general forecasting objective of Delphi. RAM was also designed to enable the measurement of over and under use of medical and surgical interventions and their appropriateness in terms of risk and benefit for managing a condition. Thus, RAM departs from Delphi in one core attribute, in that it allows F2F meetings, arguing that many of the concerns regarding undue influence introduced by one expert to another can be ameliorated through a competent moderator for the meeting. RAM utilises a systematic literature review as the basis of the exercise, upon which expert consensus can be generated to bridge gaps in the evidence.

Though the RAM manual purports to observe consensus, it includes feedback to experts on the responses to the previous rounds. RAM also enables experts to modify the list on which the consensus is being gained by focussing on areas of disagreement, which is a key mechanism by which consensus is forced, leading to the introduction of structural bias. The RAND institute has not published any further modifications to Delphi other than RAM; therefore, studies not meeting the description of RAM are assigned to the category entitled nRMD. These are studies where the methodology has been designed and described by the authors themselves.

Consensus generating methodologies are designed to answer a research question and contain two broad stages. Firstly, there is an item generation phase, during which the subject of the consensus is defined, and a list of answers to the research question is solicited from the experts and/or independent sources. During the Second phase, also known as a consensus phase, the list of items generated in answer to the question is provided to a group of experts, who then provide their opinion on the content of the list. Consecutive consensus rounds are used to narrow down the experts' opinion using a threshold for agreement, to arrive at the final constituents of the list.

To reflect these, three key characteristics were selected to assess the homogeneity of the population of the three methods: item generation phase, consensus phase, and consensus assessment criteria. Tests were carried out to assess both inter- and intra-group homogeneity using a student's *t*-test ($p < .05$) and a variation threshold of 10% for each characteristic, respectively.

Forced versus observed consensus (control and study methodologies)

The Jandhyala method advocates a passive or observational approach to consensus measurement[12]. The premise for this is that methods employing a consensus-forcing approach risk the integrity of the final consensus, through actively influencing the initial level of agreement observed, thereby introducing a structural bias to the consensus generating process.

For the purposes of this research, the act of forcing a consensus is defined as any procedure that alters the composition of a list (through either adding or removing items) on which consensus had already been sought, for it to be resubmitted for a further consensus round, thereby excluding dissenting opinion to the changes and encouraging experts to "change" their minds on the list's composition. By contrast, observing a consensus is achieved by preserving the initial item list and recording experts' level of consensus on the inclusion of each item in that list; dissenting opinions can be observed unchanged *via* item ranking and predefined consensus thresholds.

To assess whether the Jandhyala method is different from the control group in its approach to observing consensus, each included study was appraised against the following characteristics of observed consensus: predefined consensus criteria, and study to be stopped by reaching a set number of rounds. Secondly, studies were also assessed against the presence of the following markers of forced consensus: if the study is to be stopped after a consensus is reached, pre-specified criteria for dropping of items between rounds, reviewers ability add to or reduce the item list between rounds, and whether there was feedback between rounds to inform experts decision to change their mind on a previous answer.

Given that the principle of observing consensus is the absence of all markers of forced consensus, for the Jandhyala method to be considered novel in this regard, all control studies needed to demonstrate at least one marker. Due to the binary nature of this approach, no further statistical tests were deemed necessary.

Assessment of awareness

The Jandhyala method includes two measures of awareness; these are, item awareness and advisor awareness. Item awareness was defined as the relative frequency of suggestion of each item as a factor of the most commonly occurring item (frequency of item/frequency of most common item), whilst advisor awareness was defined as the overall frequency of each item as a factor of the total number of experts (item frequency/total number of experts). The collection of these data informed the understanding of which items are at the forefront of the experts' minds and how widespread awareness of these items is across the group of experts. The calculation of an awareness index for each item enables a measure of "prompting" to be observed for each item between the item generation rounds and consensus round, by comparing the difference between this and the consensus index calculated per item in that round.

Awareness can also infer how knowledgeable experts are on the subject matter. All included studies were examined to assess whether they collected and reported on the level of item and/or advisor awareness. Furthermore, to assess whether a proxy measure of awareness could be calculated for each of the control studies, studies were assessed to determine whether experts were involved in the item generation phase of the study, alongside whether the initial item lists were generated independently (expert assessment and literature review conducted in parallel), or whether experts could comment on the pre-determined list of initial items. Given the binary nature of this approach, no further statistical tests were deemed necessary.

Minimum expert engagement profile

The Jandhyala method adopts a strict two-round approach to arrive at an observed consensus, with one item generation round and one consensus round. Furthermore, it does not utilise a F2F meeting at any stage in its execution; it is therefore denoted as utilising a $1+1+0$ design. This design is intended to both preserve anonymity of the experts, ensuring each contributes equally to both awareness and consensus measures, and enable a rapid arrival at a consensus with the minimum engagement possible. The avoidance of any logistical challenges associated with arranging meetings in order to achieve the consensus is also a perceivable benefit.

For the Jandhyala method to demonstrate differentiation and uniqueness in this profile, no other studies in the control group should exhibit the same profile. The systematic review results were interrogated against the following criteria: the number of publications where a single item generation round and single consensus round were used, and the number of F2F meetings were zero ($1+1+0$ profile). Given the binary assignment, no further statistical tests were deemed necessary.

Efficiency of the minimum expert engagement profile

The minimum ($1+1+0$) expert engagement profile was purported to be quicker than the control studies. The uniqueness of this minimum achievable profile in the broader population of consensus generating approaches can infer both improvements in speed and in the resource. However, an assessment of time duration would also be of benefit. Duration of each method ($1+1+0$ vs non-$1+1+0$) was collected, and a student's *t*-test ($p<.05$) was used to detect any statistically significant differences between the means. The use of this minimum engagement may also serve to maintain expert interest in the exercise, thereby reducing the attrition between subsequent rounds. Attrition rates were extracted where possible from study designs, and any difference in means was tested with a student's *t*-test ($p<.05$).

Results

The initial search yielded a total of 831 results through MEDLINE *via* PubMed and 350 results through Google Scholar. There were 961 results in total identified for screening after duplicates were removed. After filtering, utilising the quality control checklist, and selecting studies based on their titles and abstracts, 886 studies were selected. From these, 75 studies were excluded for not meeting the inclusion criteria, and another 47 were excluded after the remaining full-text articles ($n=811$) were assessed for eligibility. Thus, the remaining 764 studies were included in the final list of studies and progressed to the full-text review stage (Figure 1; Supplements 1 and 2).

Delphi and derivatives versus Jandhyala method

Study methodology type

Overall, from the 764 Delphi and Delphi derivative studies included in this analysis:

- 329 (43.06%) were identified as using traditional Delphi methodology
- 426 (55.76%) were identified as using nRMD methodologies
- 9 (1.18%) were identified as using the RAM

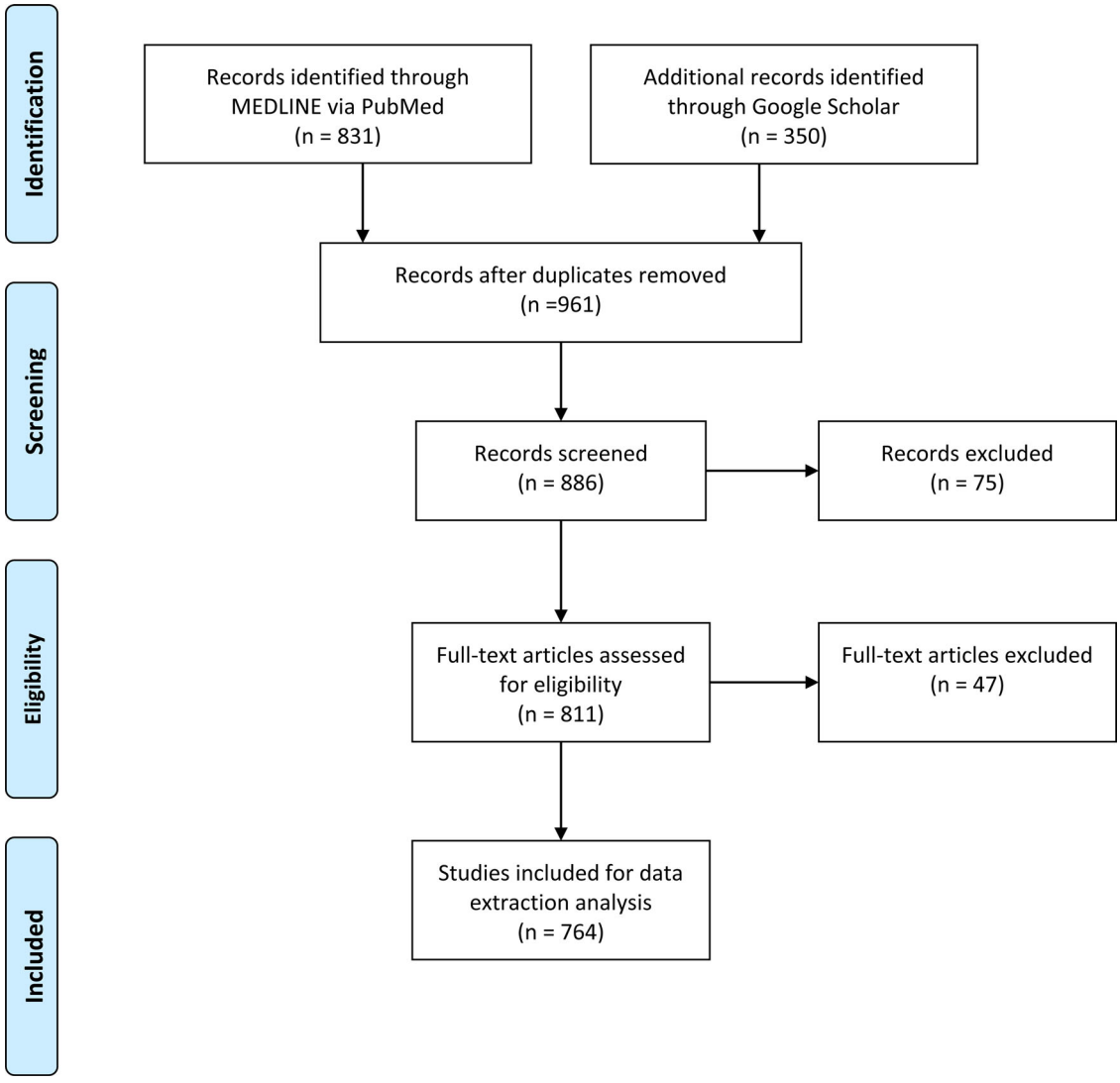

Figure 1. PRISMA flow diagram – Page 10.

Four instances of the Jandhyala method were included in the comparator assessment and analysis.

Heterogeneity of Delphi, nRMD and RAM methods

Heterogeneity in the item generation phase was observed across all examined characteristics in the three Delphi-based methods, with the exception of the inclusion of a distinct pre-survey panel ($p = .667$). Of the five characteristics in the item generation phase (excluding the number of rounds), intra-method variation above the 10% margin was observed for 4/5 (80%) Delphi, 4/5 (80%) nRMD and 1/5 (20%) RAM methodologies.

39/329 (11.89%) Delphi studies used F2F meetings between the experts, with 105/329 (32.01%) using a systematic literature review alongside experts in the item generation phase and 34/329 (25.60%) using only experts.

Inter-method heterogeneity in the consensus phase was observed across all measured characteristics in all three Delphi-based methods, with the exception of the number of rounds ($p = .14$). In this phase, Delphi studies were observed to have used F2F meetings in 64/329 (19.51%), a practice that was discouraged in the original principles of Delphi-based methods.

Of the three characteristics in the item generation phase (excluding the number of rounds) which comprised of only F2F meetings (before, during and after consensus rounds), intra-method variation above the 10% margin was observed for 2/3 (66.67%) Delphi, 3/3 (100%) nRMD and 2/3 (66.67%) RAM methodologies.

Of the *consensus assessment* criteria, inter-method heterogeneity was observed for 2/5 measures – RAND/UCLA criteria ($p < .001$) and item ranking ($p = .038$). Of these, RAND/UCLA criteria were synonymous with RAM. However, these similarities were only observed in 6/9 (66.67%) studies purporting to describe themselves as RAM.

Finally, no inter-group heterogeneity was observed for any Delphi-derived methodologies for either of the two stopping rules: stop by consensus ($p = .165$) and stop by the number of rounds ($p = .082$). All observations 6/6 (100%) were above the intra-method threshold for heterogeneity. A summary of this data can be found in Table 3.

It is important to note that, the heterogeneity in the methodology characteristics between the Delphi, RAM and

Table 3. Showing heterogeneity in item generation phase, consensus phase, and consensus assessment criteria across Delphi, nRMD and RAM (Page 11).

	Delphi (329)	nRMD (426)	RAM (9)	*p*-Value
Heterogeneity in item generation phase				
Number of rounds *N* (%) (lowest, Q1, median, Q3, highest)	(0,1,1,1,9)	(0,1,1,2,8)	(0,1,1,1,4)	<.001
F2F meetings (Yes, %)	39 (11.89)*	84 (19.72)#	**0**	.006
Method of item generation				
Both Expert survey and SLR *N* (%)	105 (32.01)	234 (54.93)#	3 (33.33)	<.001
Only SLR *N* (%)	84 (25.60)*	113 (26.53)	5 (55.56)	
Only Expert survey *N* (%)	133 (40.42)	73 (17.14)	1 (11.11)	
Neither Expert survey nor SLR *N* (%)	**7 (2.13)**	**6 (1.41)**	**0**	
Item list generated independently (Yes, %)	185 (56.40)	201 (47.18)	5 (55.56)	.006
Experts impacting final list (Yes, %)	159 (48.48)#	259 (60.80)#	7 (77.78)*	<.001
Distinct pre-survey panel (Yes, %)	187 (56.83)#	240 (56.34)#	7 (77.78)	.667
Heterogeneity in consensus phase				
Number of rounds *N* (%) (lowest, Q1, median, Q3, highest)	(1,2,3,3,7)	(1,2,3,3,10)	(2,2,2,3,4)	.14
F2F meetings				
Before (Yes, %)	39 (11.89)*	84 (19.72)#	**0**	.006
During (Yes, %)	**25 (7.62)**	69 (16.08)#	2 (22.22)#	.001
After (Yes, %)	34 (10.36)*	76 (17.84)#	2 (22.22)#	.012
Any F2F meetings (Yes, %)	64 (19.51)*	135 (38.73)#	4 (44.44)#	<.001
Heterogeneity in consensus assessment criteria and stopping criteria				
Percentage consensus *N* (%)	262 (79.64)	320 (75.12)	6 (66.67)	.129
Mean and/or Median scores *N* (%)	140 (42.68)	163 (38.26)	5 (55.56)	.319
Inter-percentile range *N* (%)	45 (13.72)	57 (13.38)	**0**	.484
Item ranking *N* (%)	68 (20.73)	114 (26.76)	**0**	.038
RAND-UCLA criteria *N* (%)	**12 (3.65)**	**38 (8.92)**	6 (66.67)*	<.001
Stopping rules				
Stop by consensus *N* (%)	148 (45.12)*	162 (38.03)#	4 (44.44)#	.165
Stop by number of rounds *N* (%)	148 (45.12)*	220 (47.42)#	3 (33.33)#	.082
Number of methodology-defining characteristics exceeding Intra-method variability threshold (*)	6	0	2	
Number of other characteristics exceeding Intra-method variability threshold (#)	2	10	5	
Total exceeding threshold among all characteristics (excluding number of rounds in item generation and consensus phases and any F2F meetings)	8/15	10/15	7/15	

Values that exceed this threshold are marked with an asterisk (*) if they form part of the definition. If they do not form part of the definition, they are marked with an (#). In addition, we have highlighted in bold font, values from the evaluation of how the Delphi, RAM and nRMD methods were used, that do not apply to the methodology used in this study.

nRMD methods was tested and the associated *p*-values are presented in Table 3. Chi-square tests were used for all categorical variables, and Kruskal-Wallis tests were used for the number of rounds. The heterogeneity within each method is evaluated using a 10% threshold of all studies using that methodology for allowing any variation. Any characteristic that exceeds this threshold is marked with an asterisk (*) if the characteristic is part of the definition and with an (**#**) otherwise.

Assessment of forced vs observed consensus

In studies using the Jandhyala method, 4/4 (100%) had predefined consensus criteria, and the study was stopped after a pre-specified number of rounds; no studies using the Jandhyala method were stopped by consensus or had any pre-specified criteria for dropping items, i.e. no items could be dropped from the item list in these studies. The experts' first opinion on each item was retained, and no attempt made to encourage a change.

For markers of observed consensus, control methodology studies (Delphi, nRMD and RAM), 523/764 (68.46%) were observed to have predefined consensus criteria whilst 159/764 (20.81%) did not have predefined consensus criteria. Furthermore, 82/764 (10.73%) did not report the information on predefined consensus criteria. Control studies were observed to have stopped by the number of rounds in 314/764 (41.10%), with 352/764 (46.07%) not stopping, and 98/764 (12.83%) not reporting this information.

For markers of forced consensus in the control group, 371/764 (48.56%) stopped by consensus, 301/764 (39.40%) did not stop by consensus, and 92/764 (12.04%) did not report this information. 600/764 (78.54%) provided feedback between rounds, 148/764 (19.37%) did not provide feedback between rounds, and 16/764 (2.09%) did not report this information.

Furthermore, 605/764 (79.18%) had pre-specified criteria for dropping items in subsequent rounds, 157/764 (20.55%) did not have pre-specified criteria for dropping items in subsequent rounds, and 2/764 (0.26%) did not report this information. 344/764 (45.03%) enabled experts to add items to the list between consensus rounds, 383/764 (50.13%) did not enable experts to add items to the list between consensus rounds, and 37/764 (4.64%) did not report this information. 497/764 (65.05%) control studies allowed experts to reduce items during consensus rounds, 229/764 (29.97%) did not allow experts to reduce items during consensus rounds, and 38(4.97) did not report this information.

Finally, all control group studies (764/764, 100%) were found to have at least one marker of forced consensus. This information is summarised in Table 4, as well as an in-depth assessment of these parameters compared across all methodologies (Delphi, nRmD, RAM and Jandhyala method) can be found in Figure 2.

Table 4. Showing markers of forced consensus (Page 13).

	Control methodologies			Jandhyala method		
	Yes	No	Not reported	Yes	No	Not reported
Observed consensus						
Pre-defined consensus criteria *N* (%)	523 (68.46)	159 (20.81)	82 (10.73)	4 (100)	0	0
Stop by number of rounds *N* (%)	314 (41.10)	352 (46.07)	98 (12.83)	4 (100)	0	0
Forced consensus						
Stop by consensus *N* (%)	371 (48.56)	301 (39.40)	92 (12.04)	0	4 (100)	0
Feedback between rounds *N* (%)	605 (79.18)	157 (20.55)	16 (2.09)	0	4 (100)	0
Pre-specified criteria for dropping items in subsequent rounds *N* (%)	605 (79.18)	157 (20.55)	2 (0.26)	0	4 (100)	0
Experts allowed to add Items during consensus rounds	344 (45.03)	383 (50.13)	37 (4.84)	0	4 (100)	
Experts allowed to reduce items during consensus rounds	497 (65.05)	229 (29.97)	38 (4.97)	0	4 (100)	
Any indicator of forced consensus	764 (100)	0	0	0	4 (100)	

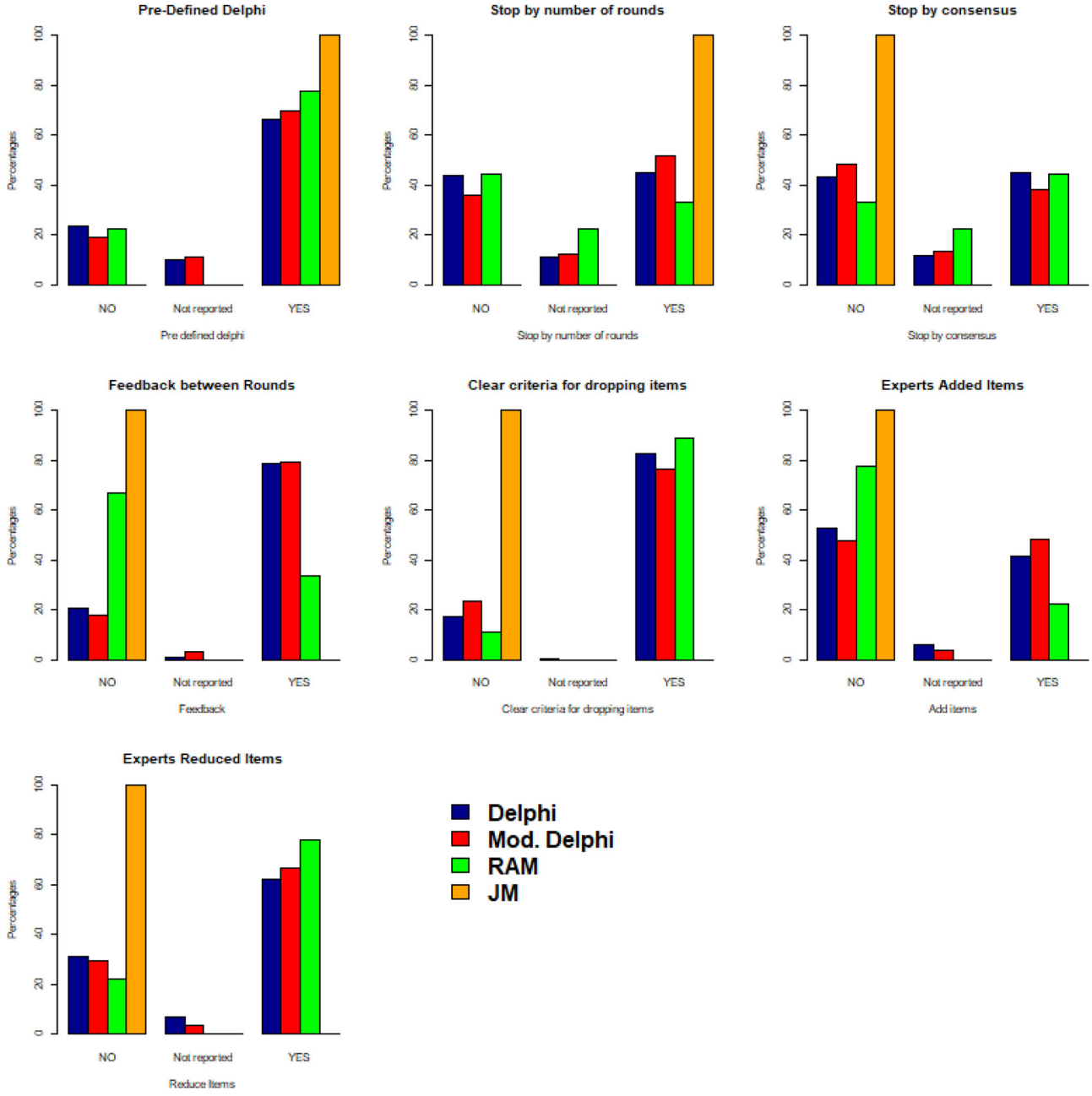

Figure 2. Forced vs observed consensus graphs – Page 14.

It is important to note that, the presence of predefined consensus criteria, and other characteristics such as, whether studies were ended following achievement of predefined consensus endpoints, whether studies were ended following a pre-specified number of rounds, and whether there were pre-specified criteria for removing items from the item list

Table 5. Showing assessment of awareness (Page 14).

	Control methodologies			Jandhyala method		
	Yes	No	Not reported	Yes	No	Not reported
Item awareness *N* (%)	0	50 (6.54)	714 (93.46)	4 (100)	0	0
Adviser awareness *N* (%)	0	47 (6.15)	717 (93.85)	4 (100)	0	0
Experts included in item generation phase *N* (%)	549 (71.86)	205 (26.83)	10 (1.36)	4 (100)	0	0
Item list generated independently *N* (%)	391 (51.18)	217 (28.40)	156 (20.42)	4 (100)		0

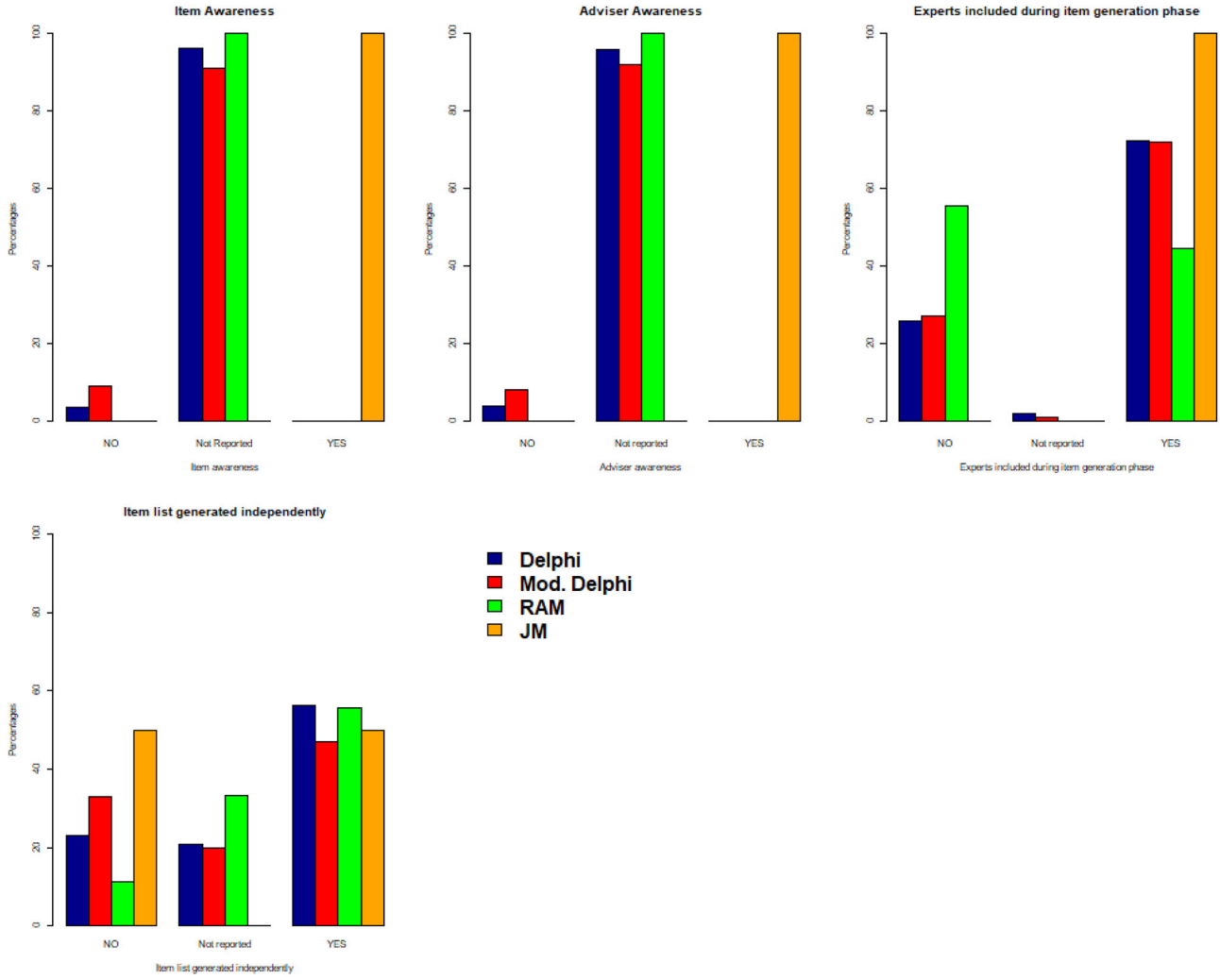

Figure 3. Awareness measures graphs – Page 15.

between rounds, were all assessed in all included studies utilising Delphi, Delphi derivative or Jandhyala methodologies.

Assessment of awareness

From the identified Delphi and Delphi derivative studies, 0/764 (0%) reported measurements of either item or advisor awareness, whilst all studies using the Jandhyala method measured and reported awareness indices 4/4 (100%).

For the assessment of proxy measures of awareness, experts were found to be included in the item generation phase for all Jandhyala method studies 4/4 (100%) and 549/764 (71.86%) control studies. They were not included in 205/764 (26.83%) control studies, and 10 (1.36%) control studies did not report on this measure. Further, 4/4 (100%) Jandhyala method studies and 391/764 (51.18%) control studies generated the initial item list independently (of the experts), whilst 216/764 (28.55%) control studies did not, and 156/764 (20.42%) did not report on this measure. A summary of these data can be found in Table 5.

It is important to note that, the inclusion of assessments of item awareness (frequency of item/frequency of most common item), advisor awareness (item frequency/total number of experts), whether experts were included in the item generation phase, and whether the item list was generated independently were assessed in all identified Delphi and Delphi derivative studies and compared against studies utilising the Jandhyala methodology (Figure 3).

Table 6. Showing the distribution of F2F meetings and number of rounds in item generation and consensus phases by Jandhyala method and control methodologies (Page 15).

	Control methodologies	Jandhyala method
F2F meetings, *N* (%)		
Yes	201 (26.30)	0
No	563 (73.69)	4 (100%)
Total number of item generation rounds		
0 round	5 (0.65)	
1 round	530 (69.37)	4 (100)
2 rounds	113 (14.79)	0
3–5 rounds	82 (10.73)	0
>6 rounds	3 (0.39)	0
Not reported	31 (4.06)	0
Total number of consensus rounds		
1 round	54 (7.07)	4 (100)
2 rounds	303 (39.66)	0
3–5 rounds	377 (49.34)	0
>6 rounds	21 (2.75)	0
Not reported	10 (1.31)	0
Total number of study rounds (item generation + consensus)		
1 round	0	0
2 rounds	39 (5.10)	4 (100)
3–5 rounds	576 (75.39)	0
>6 rounds	111 (14.53)	0
Not reported	38 (4.97)	0
Total number of studies with (1 + 1 + 0) profile		
Yes	13 (1.70)	4 (100)
No	751 (98.30)	0

Minimum expert engagement profile (1 + 1 + 0)

F2F meetings were observed in 201 (26.30%) control studies, whilst no studies utilising the Jandhyala methodology employed a F2F meeting in the study protocols (Table 6).

With respect to the number of study rounds, all studies utilising the Jandhyala methodology had one item generation phase (4, 100%), one consensus round (4, 100%) and a total of two overall study rounds (4, 100%). The majority of control studies had only one item generation round (530/764, 69.37%), with 113/764 (14.79%) having two rounds, 82/764 (10.73%) having between three and five rounds, and 3/764 (0.39%) having six or more rounds; 32/764 (4.19%) studies did not report the number of item generation rounds. For the number of consensus rounds, 54/764 (7.07%) control studies had one consensus round, 303/764 (39.66%) had two consensus rounds, 377/764 (49.34%) had between three and five rounds, and 21/764 (2.75%) had over six consensus rounds; 10/764 (1.31%) did not report the number of consensus rounds (Figure 4).

The number of studies utilising the minimum engagement profile ("1 + 1 + 0") in the control group was 13/764 (1.70%), compared to 4/4 (100.00%) with the Jandhyala method. These data are summarised in Table 6.

Minimum expert engagement profile (efficiency)

For attrition, 4/4 (100%) Jandhyala method studies (1 + 1 + 0 profile by definition) showed 0 attrition compared to 10/764 (76.92%) control studies utilising a 1 + 1 + 0 design and 283/764 (37.68%) control studies not utilising a 1 + 1 + 0 design. The difference in attrition between the Jandhyala method and control 1 + 1 + 0 designs was not statistically significant ($p = .541$), however the difference in attrition between the combined Jandhyala plus control 1 + 1 + 0 studies and the non 1 + 1 + 0 control studies was statistically significant ($p < .001$) (Figure 5).

The 1 + 1 + 0 methods were complete in 0–183 days by 4/4 (100%) Jandhyala method, and 3/764 (23.08%) control 1 + 1 + 0 methods. 1 (7.69%) control 1 + 1 + 0 study and 132 (17.58%) non-1 + 1 + 0 control studies were completed in >183 days. Studies where duration data were not reported were as follows: Jandhyala − 0/4 (0%), Control 1 + 1 + 0 − 9/764 (69.23%), and Control non-1 + 1 + 0 − 479 (63.78%). The difference in duration between the 1 + 1 + 0 design group (Jandhyala and Control) and non-1 + 1 + 0 studies was not statistically significant ($p = .999$). These results are summarised in Table 7.

Discussion

Heterogeneity of Delphi, nRMD and RAM

There were two main objectives of this study. The first objective was to define the control group, or current standard consensus generating methods (Delphi, RAM and nRMD), and ascertain if any heterogeneity exits within and between each method. The second objective was to conclude whether the newly developed Jandhyala method is sufficiently differentiated from the current standard, to constitute a discrete approach to generating expert opinion. In total, 764 control studies were identified which met the study inclusion and exclusion criteria. Of these 764 studies, 329 utilised traditional Delphi methods, 426 used nRMD methods and 9 used RAM. These 764 studies were compared against the four instances of the Jandhyala method, which were used as case studies in the method paper publication[12].

As the Jandhyala method was designed specifically to address many of the perceived limitations associated with Delphi and Delphi derivative methodologies, variables

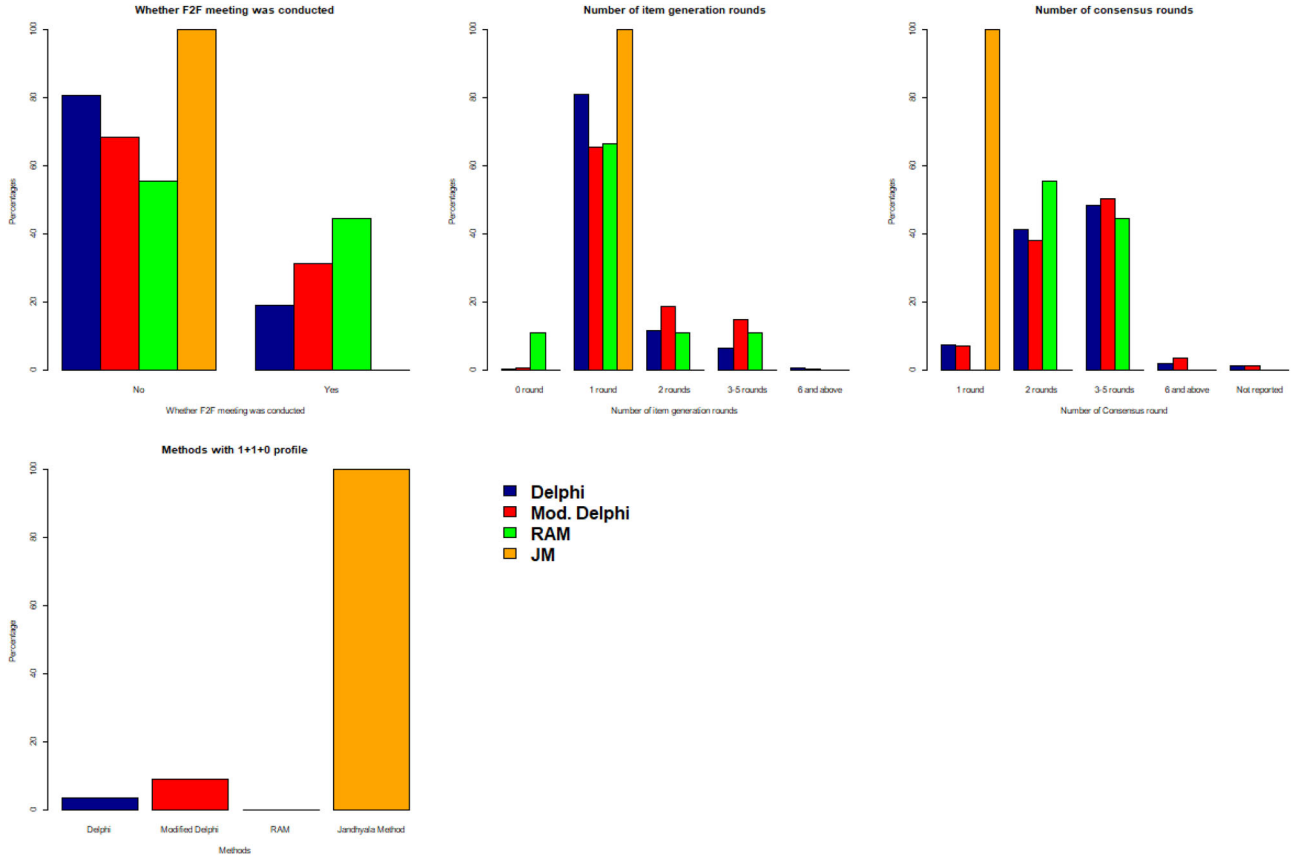

Figure 4. Study composition (rounds) graphs – Page 15.

associated with these characteristics were the primary focus of the systematic literature review, variables such as, dropping or adding items between consensus rounds, measurement of prior awareness, methods of consensus measurement, study duration and expert attrition rates, have all been previously identified as areas of concern for Delphi-style studies[9,14].

On reviewing the current "standard" consensus generating methods which constituted the control group, several findings were of interest. There was heterogeneity observed both within and between groups for the majority of the selected parameters. Several core characteristics of Delphi and RAM methodologies also showed significant intra-method, heterogeneity implying that subsequent authors were varying the original methodology to meet their immediate needs. This phenomenon is particularly apparent in the nRMD group, where no original or consistent "modification" could be identified between studies. Given the wide variation in the current standard methods, no clear individual method can be identified to act as a control. Therefore, the overall population of the three methods was used as a control group against which the Jandhyala method was assessed.

Assessment of forced or observed consensus

The differentiation between methodologies on the type of consensus they achieve – i.e. whether it was an artificially induced, "forced" consensus, or a passive, observed consensus – is fundamental to this exercise. Forced consensus is demonstrated by Delphi and its derivatives as a result of the change which they enact between rounds, in the composition of the list on which consensus is being sought. This change happens in two ways; firstly, where threshold levels of consensus on inclusion of an item are not met, these items can be dropped from the list of answers taken to the subsequent round and alternative items may be added. Secondly, the use of F2F meetings in the Delphi derivatives may allow more dominant members of the expert panel to have a disproportionate influence on the composition of the items in the list. In both situations, the result is the undermining of the dissenting opinion and acceptance of a version that has a favoured majority. It is, therefore, a more limited version of the original consensus with no reflection of its true variation.

Observed consensus, as described in the Jandhyala method, preserves the totality of the experts' initial list composition, and precludes any addition and/or subtraction before proceeding directly to the solicitation of experts' consensus on the items to be included in the complete list. This method is therefore limited to a maximum of two rounds and does not use any F2F meetings. The observed item consensus is measured by a predefined threshold consensus index (>50%). Through preserving all items above and below this cutoff, relative levels of agreement on each item are categorised, enabling all items to be handled in different ways later (adopt, adopt or investigate, investigate and monitor).

In this study, the type of consensus, forced or observed, was assessed *via* recording the method by

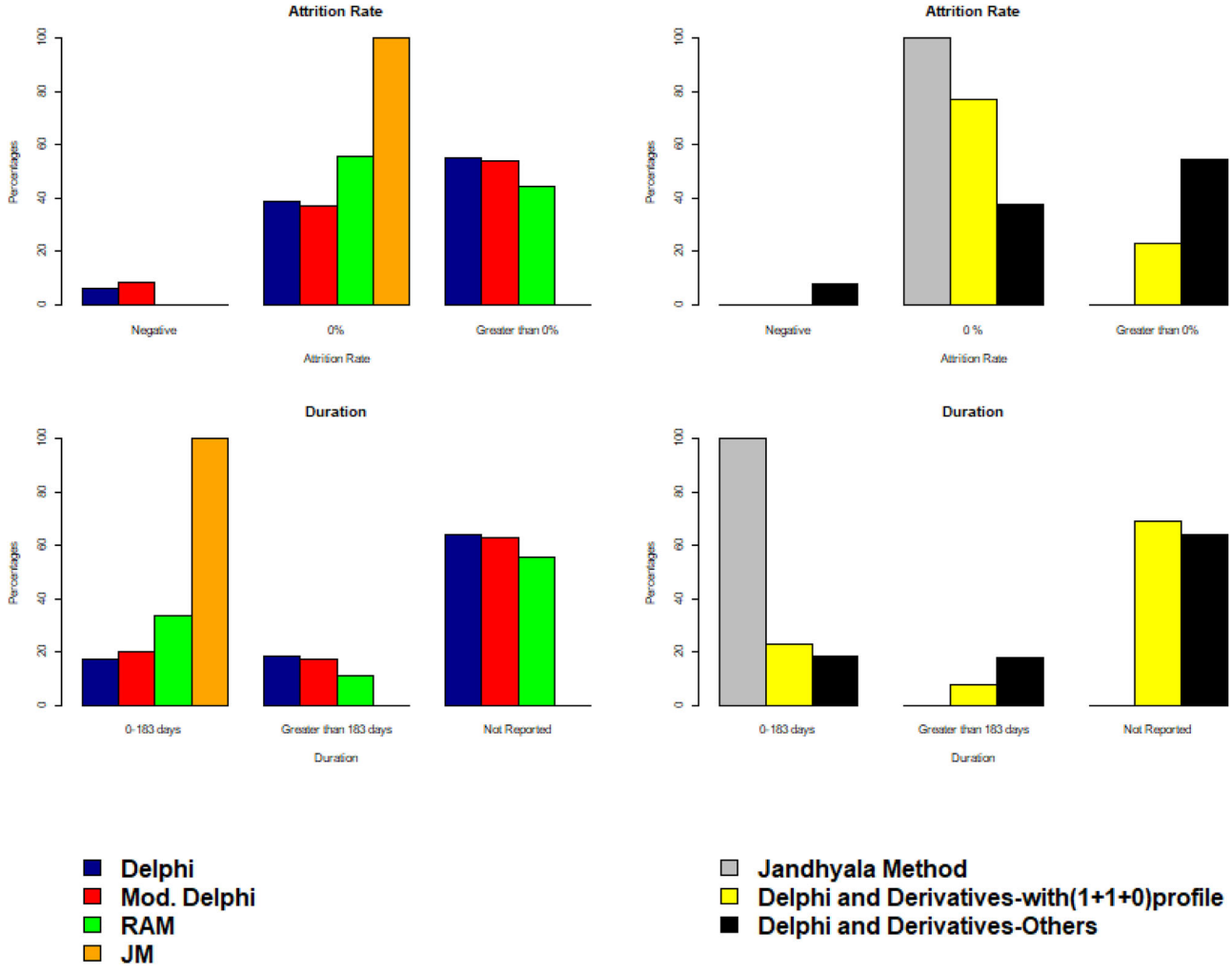

Figure 5. Efficiency, attrition and duration graphs – Page 16.

Table 7. Distribution of study attrition rate and study duration by Jandhyala method, control and non-control profiles along with the associated *p*-values for between groups comparisons (Page 16).

	Jandhyala method (JM)	Control (1 + 1 + 0) profile	Control non (1 + 1 + 0) profile
Attrition, *N* (%)			
0%	4 (100%)	10 (76.92)	283 (37.68)
>0%	0	3 (23.08)	411 (54.73)
Not reported	0	0	0
negative	0	0	57(7.59)
Comparison of 1 + 1 + 0 profile (Jandhyala method + control) vs non 1 + 1 + 0 – *p*-value			.001
Comparison of Jandhyala method vs control 1,1,0 profile – *p*-value		.541	
Duration, *N* (%)			
0–183 days	4 (100%)	3 (23.08)	140 (18.64)
>183 days	0	1 (7.69)	132 (17.58)
Not reported	0	9 (69.23)	479 (63.78)
Comparison of 1 + 1 + 0 profile (Jandhyala method + control) vs others – *p*-value			.985
Comparison of Jandhyala method vs control 1 + 1 + 0 profile – *p*-value		.999	

which consensus was measured (percentage, mean/median, range, ranking and RAND/UCLA criteria), alongside the presence or absence of any markers of a forced consensus (study stopping criteria, pre-definition of consensus criteria and criteria for dropping items between rounds listed).

The Jandhyala method and a proportion of the control methodologies group shared characteristics by which consensus on items could be observed, with 523/764 (68.46%) control studies using a predefined consensus threshold and 314/764 (41.10%) stopping by the number of rounds. This indicates a degree of mutual recognition of this approach to measuring consensus. However, whilst this remained a stereotypical part of the Jandhyala method with 4/4(100%) using this approach, it was found to be an area of heterogeneity in the control group.

When further considering markers of forced consensus, the individual incidences in the control group continue to reflect heterogeneity and lack of consistency, in the way their constituent methods have been applied. When all indicators were applied to the control group to ascertain whether any fulfilled the definition of observing a consensus, no results were returned. The Jandhyala method therefore stood differentiated from the control group in the specific manner by which consensus is achieved.

A future application of the Jandhyala method would centre around leveraging the standardised reporting of the results and the comparability of them when applying the method serially over time for the same research question, thereby enabling the accurate monitoring of changes in awareness and consensus as a result of a particular intervention.

Assessment of awareness

Measuring each expert's prior awareness of each item arising from the group's answers to the research question informs an understanding of any differentials that may exist in the minds of the experts. This has the benefit of being able to handle groups of items with different awareness levels in different ways, for example, educating on items with comparatively low awareness scores, and accepting items with complete awareness. It was observed that only the Jandhyala method measured and reported on item awareness (how frequently an item was suggested compared to the most frequent item), and advisor awareness (number of experts mentioning an item).

Whilst studies utilising Delphi and its derivatives often included experts in the item generation phases of their studies, and over half used both expert opinion and other methods, such as systematic literature reviews, which would allow for some calculation of awareness, no studies reported any information on either item or advisor awareness. This lack of reporting thus precludes – amongst other things – the ability to assess the quality of the experts recruited to the research through their unprompted knowledge of the subject matter. The Jandhyala method stands differentiated from the control group in the measurement of item and advisor awareness.

Minimum expert engagement profile

Other potential concerns which are often raised with traditional Delphi and Delphi derivative methodologies are the length of the process and the expert attrition rate throughout the study. It has been hypothesised that as study duration increases, so does expert attrition. Hence, the number of study rounds and F2F meetings, overall study duration, and the level of attrition across the study were either extracted or calculated based upon the information contained within each publication.

The Jandhyala method describes a minimal engagement strategy with its subjects, employing a single item generation round, a single consensus round, and no F2F meetings. It was found that whilst the majority of the Delphi and Delphi derivative studies had only one item generation round prior to the consensus survey rounds, consistent with the Jandhyala method studies, the total number of consensus rounds varied dramatically, with the majority of Delphi and Delphi-derivatives having between three and five consensus rounds, compared with one in the Jandhyala case studies.

When the total number of rounds was calculated (item generation rounds + consensus rounds + any F2F meetings), studies utilising the Jandhyala method remained at two rounds and no F2F meetings (or within the minimum "1 + 1 + 0" profile), whilst the majority of Delphi and Delphi-derivative studies were observed to have between three and five rounds, with a quarter of all studies having up to ten rounds, and some studies described more than eleven rounds of consensus. Importantly, 13/764 (1.70%) of the control group were observed to have also used the 1 + 1 + 0 profile, indicating that the profile was not unique to the Jandhyala method but an extremely rare occurrence in the control group.

Efficiency of minimum engagement profile

Having observed the frequencies of the minimum engagement profile in the study and control groups, the final question this study answer is, whether the study design profile (a key characteristic of the Jandhyala method) offers any efficiencies in shortening the duration of the studies and/or reducing subject attrition rates.

Interestingly, the study durations appeared evenly distributed across the selected time categories for all study types, despite the different number of study rounds. However, most studies in the control group did not report either study duration or start and end dates to allow its calculation, suggesting that, these results should be interpreted with caution. However, based on emerging evidence from this study, it can be inferred that, the greater the number of rounds involved in a study, the larger the resource burden on the researchers and study participants. For example, although several logistical considerations will need to be considered where a F2F meeting is required by the methodology, these logistical approaches will be excluded by Delphi design methods not requiring F2F meetings. Hence, the Jandhyala method, alongside any instances of control group studies utilising the 1 + 1 + 0 design profile, may be considered as utilising the minimum resource possible in achieving a consensus.

Lastly, attrition rates were assessed across all study designs. Jandhyala studies did not suffer from any attrition across the study duration, whereas around 2/3 of control group studies experienced some level of attrition. Surprisingly, except for studies with >50% attrition, all attrition rate groupings had similar frequency rates regardless of control group study type. Some Delphi and modified Delphi studies were seen to have a negative attrition rate, i.e. the study ended with more experts than it started with. Both loss and gain of experts over the study course are causes for concern when considering the integrity of the final consensus reached, as it means that not all experts have had an

equal opportunity to provide their opinions on each item, particularly if items can be removed between rounds as discussed previously.

Another consideration which is not directly related to any single variable is the high proportion of missing information in the Delphi and Delphi derivative studies. Whilst some information reporting will be different between the Jandhyala and Delphi-style studies (such as the presence of bespoke awareness outputs in the Jandhyala method), a small proportion of the Delphi and Delphi derivative studies failed to report (either specifically, or providing the information to assess independently) data on variables such as attrition and study duration, alongside methodological concepts such as pre-definition of consensus criteria. The heterogeneity observed across the control studies appears to extend into their reporting; this is particularly apparent in the nRMD group.

By contrast, the Jandhyala method has standardised reporting built into the method, allowing both consistent intra-study data interpretation and inter-study comparisons. This is an important consideration, as missing data and information from these reports do not allow accurate assessment of the stringency and reliability of the methods and results. It may also limit the interpretation and generalisability of any consensus results obtained.

Limitations

There are some limitations to this study. Firstly, due to the volume of studies which were identified in the systematic literature review, alongside the number of variables which were extracted from each included study, a team of eight researchers were involved in the data extraction step. As described, researchers were split into pairs and each pair reviewed, in duplicate, a proportion of the total number of included studies; not all studies were therefore assessed by all participating researchers, meaning that there may have been differences in data extraction methods dependent upon each individuals background and training.

However, all researchers involved in data extraction had equivalent scientific backgrounds and were trained in the use of the study extraction datasheet. A third, independent, senior researcher was also available to resolve queries raised in relation to the extracted data between the primary and secondary researchers. These measures are expected to have minimised any potential bias arising from multiple experts reviewing a section of the overall dataset.

Further, a total of 764 Delphi and Delphi derivative studies were identified and included in this study but were compared to only four instances of the Jandhyala method, all of which were conducted by a team trained in its implementation and with the involvement of its author and inventor. Importantly, given the recent publication of the study method, it remained unmodified and true to its original design. However, it would be of interest to repeat this study when more projects have been conducted using the Jandhyala method, particularly by groups who have not been directly involved in its conception and development, to assess whether a modification has been deemed necessary by subsequent researchers.

Importantly, these limitations reveal that further studies utilising the Jandhyala Method are necessary, especially if it is to be widely adopted within the scientific community; thus, there are currently several studies underway to further the knowledge and use of the method.

Conclusions

In conclusion, this study has shown that the Jandhyala method for generating group consensus and awareness is unique in observing consensus and measuring awareness of subject matter across expert experts. Though the Jandhyala method is recognised as utilising the minimum engagement profile in all its case studies to reach its consensus endpoint, this profile was nevertheless also observed in a very small number of control studies. This minimum profile is attractive in that it appears to be significantly better than the other more protracted profiles is achieving zero attrition during studies.

The Jandhyala method is novel in its approach to observing group consensus and can be used to measure expert opinion, improving upon the traditional Delphi-style methodologies through the introduction of new insights into awareness of subject matter in the expert group. This method also consistently employs the minimal engagement profile, which preserves subject participation and the integrity of the resultant consensus. However, more experience of the study protocol across a wider group of researchers is required to corroborate findings from this research.

Transparency

Declaration of funding

Medical writing and editorial support were funded by Medialis Ltd.

Declaration of financial/other relationships

RJ is the founder of Medialis Ltd. There is no publication bias or conflict of interest in the publishing of this work. Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Author contributions

RJ is responsible for work conducted and reported in this literature review and has approved the submitted version, RJ takes personal and professional responsibility for the accuracy and integrity of the work reported.

Acknowledgements

RJ thanks the medical writers at Medialis Ltd that contributed to editing the manuscript.

Data availability statement

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

ORCID

Ravi Jandhyala http://orcid.org/0000-0002-7241-7476

References

[1] Bartlett PC, Judge LJ. The role of epidemiology in public health. Rev Sci Tech. 1997;16(2):331–336.
[2] Jewell NP. Natural history of diseases: statistical designs and issues. Clin Pharmacol Ther. 2016;100(4):353–361.
[3] Alomar M, Palaian S, Al-Tabakha MM. Pharmacovigilance in perspective: drug withdrawals, data mining and policy implications. F1000Res. 2019;8:2109..
[4] Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ. 1999;318(7182):527–530.
[5] Desai VS, Camp CL, Krych AJ, et al. What is the hierarchy of clinical evidence? In: Musahl V, Karlsson J, Hirschmann MT, Ayeni OR, Marx RG, Koh JL, editors. Basic methods handbook for clinical orthopaedic research: a practical guide and case based research approach. Berlin, Heidelberg: Springer; 2019. p. 11–22.
[6] Williams BA. Perils of evidence-based medicine. Perspect Biol Med. 2010;53(1):106–120.
[7] Bourrée F, Michel P, Salmi LR. Consensus methods: review of original methods and their main alternatives used in public health. Rev Epidemiol Sante Publique. 2008;56(6):415–423.
[8] Fitch K, editor. The Rand/UCLA appropriateness method user's manual. Santa Monica: Rand; 2001.
[9] Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. Int J Nurs Stud. 2001;38(2):195–200.
[10] Dalkey N, Helmer O. An experimental application of the DELPHI method to the use of experts. Manag Sci. 1963;9(3):458–467.
[11] RAND Corporation. Delphi method. n.d. [cited 2020 Apr 16]. https://www.rand.org/topics/delphi-method.html
[12] Jandhyala R. A novel method for observing proportional group awareness and consensus of items arising from list-generating questioning. Curr Med Res Opin. 2020;36(5):883–893.
[13] Haddaway NR, Collins AM, Coughlin D, et al. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. PLOS One. 2015;10(9):e0138237.
[14] Hinz A, Michalski D, Schwarz R, et al. The acquiescence effect in responding to a questionnaire. GMS Psycho-Soc Med. 2007;4.

Appendix 1. Items extracted from each study

General Characteristics

- *Country of Conduct*
- *Country of Publication*
- *Medical Discipline*
- *Objective Type*

Initial Item List Generation

- *From literature review*
- *From panel opinion*
  - *Panel opinion based on experience*
  - *Panel opinion based on hard evidence*
- *Thematic analysis of item list by study analyst*
- *List generated independently*
- *Item awareness measurements*
- *Advisor awareness measurement*
- *Experts can impact initial list of items prior to survey*
  - *Reviewed and impacted by consensus*
  - *Reviewed and impacted by ranking*
- *Survey type*
  - *Online survey*
  - *Postal survey*
  - *Process executed via face-to-face discussions*
- *Total number of items generated pre-survey*

Literature Review

- *Literature review design*
  - *Structured/systematic*
  - *Unstructured*
- *Number of medical databases searched*
- *Were guidelines reviewed*
- *Were speciality websites or sources searched*
- *Number of analysts in the literature review*

Panel Characteristics

- *Distinct pre-survey panel*
- *Selection criteria specified (item generation phase)*
- *Number of experts/experts*
- *How many experts proceeded to survey round*
- *Country/Region of Survey experts*
- *Selection criteria specific (survey experts)*

Consensus

- *Definition*
- *Consensus method*
  - *Consensus by percentage agreement*
  - *Consensus by Mean/Median score*
  - *Consensus by inter-percentile range*
  - *Consensus by Rank*
  - *Consensus by RAND/UCLA criteria*
  - *Consensus by stability of responses*
  - *Other consensus method*
  - *Defined pre-Delphi*
- *Total number of rounds (author stated)*
- *Number of rounds for consensus*
- *Stopping criteria*
  - *Stop by consensus*
  - *Stop by number of rounds*
  - *Measurement of item performance (different between item generation and survey phases)*
  - *Other criteria*

Survey Overview

- *Survey delivery method*
  - *Face-to-face meeting conducted*
    - *Face-to-face meeting before first survey round*
    - *Face-to-face meeting between survey rounds*
    - *Face-to-face meeting after last survey round*
    - *Face-to-face meeting (sequence unclear)*
  - *Electronic*
  - *Postal questionnaire*
  - *Not reported*
- *Type of responses*
  - *Open-ended*
    - *Justification to response*
    - *Feedback/comments*
    - *Additional item generation*
  - *Rating scales*
    - *Rating on scale*
      - *Rating via Likert Scale (min)*
      - *Rating via Likert Scale (max)*
      - *Other scale (name)*
      - *Other scale (min)*
      - *Other scale (max)*
  - *Other responses*
    - *Voting*

        - *Ranking/prioritisation*
        - *Review or approval of final framework*
        - *Other (name)*
- *Impact of survey experts*
    - *Add items*
    - *Reduce items*
- *Attrition of items*
    - *Non-attrition of items/terms*
- *Type of feedback provided by investigator*
    - *Individual feedback*
    - *Group feedback – qualitative*
    - *Group feedback – quantitative*
    - *Supplementary facts*
    - *List of items newly generated*
- *Total items generated at the end of the survey*
- *Duration*
    - *Start Date*
    - *End Date*
- *Retention/Attrition*

    - *Number of experts contacted*
    - *Number of experts agreed to participate*
    - *Number of experts at the end of the survey*

Survey Round One

- *Corresponding Round Number*
- *Review of Items by Experts*
    - *Review by consensus*
    - *Review by ranking/rating*
    - *Review by face-to-face discussion*
- *Process executed via:*
    - *Online survey*
    - *Postal survey*
    - *Face-to-face discussion*
- *Number of experts involved*
- *Reviewers add items*
- *Reviewers reduce items*
- *More Delphi rounds*

(Repeat for total number of rounds per study, max identified = 8)

Retention and Attrition Rates

- *Retention Rate*
    - *Round 1*
        - *(number of experts involved in R1/number of experts agreed to participate in study)*
    - *Round 2*
        - *(number of experts involved in R2/number of experts involved in R1)*
    - *Round 3*
        - *(number of experts involved in R3/number of experts involved in R2)*
    - *Round 4*
        - *(number of experts involved in R4/number of experts involved in R3)*
    - *Round 5*
        - *(number of experts involved in R5/number of experts involved in R4)*
    - *Round 6*
        - *(number of experts involved in R6/number of experts involved in R5)*
    - *Round 7*
        - *(number of experts involved in R7/number of experts involved in R6)*
    - *Round 8*
        - *(number of experts involved in R8/number of experts involved in R7)*
    - *Round 9*
        - *(number of experts involved in R9/number of experts involved in R8)*
    - *Round 10*
        - *(number of experts involved in R10/number of experts involved in R9)*
- *Attrition Rate*
    - *(number of experts involved in R1) – (Number of experts at the end of the survey/Number of experts involved in R1)*

Quality Reporting

- *Well defined objective*
- *Rationale for using Delphi*
- *Clear justification for selection of experts*
- *Clear criteria for reducing items*
- *Clear description of study methods*
- *Presence of a flow chart/diagram*
- *Clear definition of consensus*
- *Pilot test of instrument*
- *Transparent reporting of results*
- *Statistical data analysis clearly reported*
- *Information on number of rounds*
- *Discussion of study limitations*
- *Adequacy of conclusions*
- *Methodology Naming*
    - *Is the author's name for the study correct*
    - *What would the analyst define the method as based upon study methods*