

Enhancing rigor in narrative reviews: protocol for the validation of a structured narrative review framework via Delphi consensus

Matteo Luigi Giuseppe Leoni,¹ Y Van Tran,² Van Phong Pham,² Giustino Varrassi^{3,4}

¹Department of Medical and Surgical Sciences and Translational Medicine, Sapienza University of Rome, Italy; ²Department of Anesthesiology, Tam Anh General Hospital, Ho Chi Minh city, Vietnam; ³Research Department, Fondazione Paolo Procacci, Rome, Italy; ⁴College of Medicine, University of Baghdad, Iraq

Abstract

Narrative reviews remain essential for synthesizing complex, multidisciplinary evidence, particularly in heterogeneous and evolving fields. However, their intrinsic subjectivity, risk of selection bias and limited reproducibility have raised important methodological concerns. This study aims to describe the development of the Structured Narrative Review (SNR) Framework, an advanced model designed to enhance methodological rigor while preserving interpretative flexibility, and to present the protocol for its formal validation through an international expert consensus process. The SNR Framework integrates established reporting standards, including SANRA, PRISMA 2020 and PRISMA-ScR. It is organized into core domains addressing reproducible search strategies, evidence stratification, mechanistic synthesis and epistemological transparency. Although the framework is conceptually grounded in methodological triangulation, formal validation is required to confirm its clarity, feasibility and applicability across disciplines. This protocol describes an international modified Delphi consensus study designed to validate the SNR Framework. The primary objective is to achieve expert agreement on the relevance, clarity, and feasibility of each core domain, as well as on the overall framework structure, standardized scoring system and recommendations for integration into editorial and peer-review workflows. The consensus process will involve two to three sequential electronic rounds conducted in accordance with CREDES and ACCORD recommendations. An international panel of 30-50 experts will independently rate each SNR domain across dimensions of relevance, clarity and feasibility. Consensus will be defined as $\geq 80\%$ agreement. Items failing to reach consensus will be revised and resubmitted in subsequent rounds. As this is a protocol, no results are yet available. Upon completion, the study is expected to yield a validated SNR Framework with expert-endorsed core domains, a standardized scoring system and consensus-based implementation recommendations, thereby advancing transparency, rigor and reproducibility in narrative review methodology.

Key words: structured narrative review; Delphi study; consensus methodology; evidence synthesis; reporting standards; SANRA; PRISMA; research methodology.

Correspondence to: Matteo Luigi Giuseppe Leoni, Department of Medical and Surgical Sciences and Translational Medicine, Sapienza University of Rome, 00189 Rome, Italy.

E-mail: matteolg.leoni@gmail.com

Introduction

Narrative reviews have historically played a central role in synthesizing scientific knowledge, particularly in fields characterized by conceptual complexity, emerging evidence and interdisciplinary integration.^{1,2} Unlike systematic reviews, which prioritize narrowly defined questions and quantitative aggregation, narrative reviews allow for interpretative synthesis, hypothesis generation and mechanistic integration across heterogeneous domains.³ This flexibility is especially relevant in areas such as pain medicine, neuroimmunology, oncology and systems biology, where evidence spans molecular, clinical and behavioral levels.

However, this flexibility has also generated persistent methodological concerns. Narrative reviews have been criticized for lack of transparency in study selection, susceptibility to author bias,

and limited reproducibility, potentially undermining their scientific credibility.⁴ In response, Baethge *et al.*⁵ introduced the Scale for the Assessment of Narrative Review Articles (SANRA), a six-item instrument designed to improve reporting quality and editorial assessment. While SANRA has been widely adopted, it remains primarily a reporting-oriented tool and does not fully operationalize the level of methodological rigor expected by contemporary evidence synthesis standards.

Over the past decade, the landscape of scientific publishing has evolved substantially. The introduction of frameworks such as PRISMA 2020,⁶ PRISMA-ScR⁷ and advances in evidence grading methodologies have established new expectations for transparency, reproducibility, and structured reporting across review types.⁸ Consequently, narrative reviews are increasingly expected to demonstrate methodological accountability without losing their distinctive integrative strengths.

To address this need, the Structured Narrative Review (SNR) Framework has been proposed as a methodological advancement designed to bridge the gap between traditional narrative and systematic approaches. The SNR framework will introduce core domains encompassing conceptual scope and framing, reproducible search strategies, study selection transparency, evidence stratification, critical appraisal, mechanistic and integrative synthesis, quantitative awareness, bias recognition and epistemological transparency, clinical and translational relevance, reference integrity, structured reporting and limitations reporting. Additionally, the framework includes optional high-level extensions for semi-systematic hybridization and interdisciplinary integration, supported by a standardized scoring system (0-24 points) to facilitate editorial evaluation and peer review. It is acknowledged that, at this stage, the boundaries between some adjacent domains remain broad and may partially overlap. Precise operational definitions, specifying the unique scope, assessment criteria and distinguishing features of each domain, will constitute a primary output of the Delphi consensus process itself, enabling expert-driven clarification and refinement before the framework is finalized. Furthermore, the equal weighting of all domains in the scoring system reflects the current absence of empirical evidence to support differential weighting; this assumption will be explicitly submitted to expert evaluation during the Delphi process, and panelists will be invited to propose alternative weighting schemes if they consider certain domains to be of disproportionate methodological importance. Although the SNR framework will be conceptually developed through rigorous methodological triangulation, integrating principles from SANRA, PRISMA 2020, PRISMA-ScR and established guidance on narrative synthesis,^{9,10} formal expert validation is essential to confirm its content validity, operational feasibility and generalizability across scientific disciplines. The Delphi technique represents a well-established and widely accepted methodology for achieving structured expert consensus in situations where empirical evidence alone is insufficient to establish definitive recommendations.¹¹⁻¹³ Modified Delphi approaches have been extensively employed in clinical medicine and health research to develop consensus statements, diagnostic criteria, clinical guidelines and quality assessment tools.¹⁴⁻¹⁶

The present protocol describes the design and methodology of an international modified Delphi consensus study aimed at validating the SNR framework. The study has been designed in accordance with the Conducting and REporting of DELphi Studies (CREDES) recommendations¹⁷ and the ACCurate COnsensus Reporting Document (ACCORD) guideline,⁸ ensuring methodological rigor, transparency and reproducibility.

Primary objectives

The primary objective of this Delphi consensus study is to achieve international expert agreement on key components of the SNR framework. Specifically, the study aims to evaluate the relevance, clarity and feasibility of each of the core SNR domains, as well as to assess the overall structure and comprehensiveness of the framework architecture. In addition, consensus will be sought on the appropriateness and applicability of the standardized scoring system (0-2 per domain) and its associated quality thresholds. Finally, the study intends to generate expert-driven recommendations for the adoption of the SNR framework within editorial processes and its integration into peer-review workflows.

Secondary objectives

Secondary objectives include identifying priorities for subsequent empirical validation studies, evaluating the need for discipline-specific adaptations, and defining pathways for dissemination and endorsement by scientific journals and professional societies.

Methods

Study design

The validation of the SNR framework will follow a structured international consensus process based on a modified Delphi methodology. The Delphi method is a systematic and iterative approach designed to obtain convergence of expert opinion through sequential rounds of anonymous evaluation and controlled feedback.¹¹ This methodology is particularly well suited when evidence is emerging, heterogeneous, or insufficient to support definitive conclusions through empirical research alone.^{12,13} The modified Delphi approach adopted in this study incorporates an initial conceptual framework developed a priori by the Steering Committee, which is then submitted to the expert panel for iterative evaluation and refinement, rather than beginning with entirely open-ended questions.¹⁴

The reporting of this consensus initiative will adhere to the CREDES recommendations¹⁷ and the ACCORD guideline,¹⁸ as well as the EQUATOR Network standards for transparent reporting of health research.¹⁹ This protocol has been developed following the principles outlined in the SPIRIT 2013 statement for clinical trial protocols, adapted for consensus methodology.¹⁹

Steering committee

A multidisciplinary Steering Committee will oversee the design and execution of the consensus process. The committee will consist of 6-10 internationally recognized experts with documented expertise in evidence synthesis methodology, clinical epidemiology, translational research and scientific publishing. The composition will ensure disciplinary diversity across relevant fields, including but not limited to clinical research methodology, biostatistics, library science and editorial practice.

The responsibilities of the Steering Committee will include: defining the conceptual framework and operational definitions for each SNR domain; preparing the Delphi survey instruments; analyzing responses and determining consensus outcomes; preparing the final consensus manuscript; and developing implementation recommendations. All members of the Steering Committee will disclose potential conflicts of interest prior to participation, consistent with the International Committee of Medical Journal Editors (ICMJE) recommendations.²⁰

Selection of the expert panel

The Delphi panel will include 30-50 international experts selected according to predefined eligibility criteria. This sample size is consistent with published recommendations for Delphi studies in health sciences, in which panels of 15-50 purposively selected participants are considered sufficient to achieve reliable consensus.²¹⁻²³ Larger panels are recommended when broad geographic

representation and disciplinary diversity are sought, as in the present study.

Potential panelists will be identified based on the following criteria: authorship of at least two peer-reviewed publications in the fields of evidence synthesis, narrative or systematic review methodology, or clinical research methodology; leadership or active membership in international organizations related to evidence-based medicine or research methodology (e.g., Cochrane, JBI, EQUATOR Network); demonstrated expertise in narrative synthesis, scoping reviews, or methodological development; and geographic representation across multiple continents to enhance the generalizability of the consensus. To operationalize disciplinary and geographic balance, the Steering Committee will apply a structured allocation strategy.

At least five major disciplinary domains will be represented: i) evidence synthesis and systematic review methodology; ii) clinical epidemiology and biostatistics; iii) scientific publishing and editorial practice; iv) library and information science; v) applied clinical research, with representation spanning at least two distinct specialty areas in which narrative reviews are routinely used to synthesize heterogeneous or emerging evidence (e.g., pain medicine, oncology, neuroimmunology, public health or translational medicine). Geographically, recruitment will target representation from at least four of the six WHO regions (Africa, the Americas, South-East Asia, Europe, Eastern Mediterranean and Western Pacific), with no single region accounting for more than 50% of the panel. The Steering Committee will prospectively track recruitment against these targets and adjust outreach strategies accordingly. A diversity report will be included in the final manuscript as recommended by CREDES.¹⁷

Experts will be invited via personalized email containing detailed information about the study objectives, methodology, expected time commitment and data handling procedures. Participation will be voluntary and informed consent will be obtained electronically prior to enrollment.

Development of the initial survey instrument

The preliminary Delphi survey instrument will be developed by the Steering Committee based on the core domains of the SNR framework. Each domain will be presented with its operational

definition, key components and assessment criteria. For each domain, panelists will be asked to evaluate the following dimensions:

1. Relevance: Does this domain address a meaningful methodological aspect of narrative review quality?
2. Clarity: Is the domain definition clear, unambiguous and operationally interpretable?
3. Feasibility: Can this domain be reliably assessed in practice by editors, reviewers and authors?

Each dimension will be rated using a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = partially agree, 4 = agree, 5 = strongly agree). Additionally, panelists will be invited to evaluate the overall framework structure, the proposed scoring system and the optional high-level extensions. Open-text fields will be provided for qualitative comments, suggested modifications and proposals for additional domains. The survey instrument will be piloted with a small group of methodological experts (n = 3-5) external to the Delphi panel prior to Round 1 to assess comprehensibility, completion time and technical functionality.

Table 1 provides an overview of the core domains of the SNR framework, outlining the key methodological components used to guide the Delphi evaluation and ensure rigor, transparency and clinical relevance in the synthesis process.

Delphi process

The consensus process will consist of two, or if required, three sequential Delphi rounds conducted electronically using a secure online survey platform (Google Forms). Anonymity of responses will be maintained throughout all rounds, consistent with the core principles of the Delphi technique.^{11,17}

Round 1

In the first round, panelists will independently evaluate each of the SNR domains across the three dimensions of relevance, clarity and feasibility using the 5-point Likert scale described above. Additionally, panelists will rate the overall framework structure, the scoring system (0–2 per domain), and the proposed quality thresholds ($\geq 80\%$ = high quality; 70–79% = near-consensus; $< 70\%$ = insufficient). Participants will also be invited to provide qualita-

Table 1. Overview of the proposed core SNR domains submitted for Delphi evaluation.

No.	Domain	Description
1	Conceptual scope and framing	Explicit definition of review boundaries using structured approaches (e.g., PCC)
2	Reproducible search strategy	Transparent reporting of databases, temporal scope, search terms, and Boolean operators
3	Study selection transparency	Explicit inclusion/exclusion criteria with clear selection logic
4	Evidence stratification	Systematic categorization of evidence by methodological rigor (RCTs, observational, preclinical)
5	Critical appraisal	Structured qualitative assessment of internal validity and methodological limitations
6	Mechanistic and integrative synthesis	Cross-domain integration linking biological mechanisms with clinical outcomes
7	Quantitative awareness	Reporting of numerical findings (effect sizes, prevalence) where available
8	Bias recognition and epistemological transparency	Explicit discussion of selection, interpretative, and publication bias
9	Clinical and translational relevance	Clear articulation of implications for practice and research
10	Reference integrity	Use of peer-reviewed primary sources with DOI verification
11	Structured reporting	Organized manuscript (Introduction, Methods, Synthesis, Discussion, Limitations, Conclusions)
12	Limitations of the narrative approach	Acknowledgment of inherent limitations of narrative synthesis

PCC, population-concept-context; RCTs, randomized controlled trials; DOI, digital object identifier.

tive comments, suggest modifications to domain definitions or scope and propose additional domains if warranted. The Steering Committee will analyze all responses, calculate descriptive statistics, and identify items reaching or failing to reach the predefined consensus threshold. Items failing to achieve consensus will be revised based on aggregated expert feedback.

Round 2

In the second round, panelists will be provided with the aggregated results from Round 1, including summary statistics (median, interquartile range, percentage agreement) and anonymized qualitative feedback from other participants. Revised items and any newly proposed domains will be presented for re-evaluation using the same Likert scale. This controlled feedback mechanism allows experts to reconsider their assessments in light of the collective group opinion while maintaining anonymity, thereby facilitating convergence toward consensus.¹¹ Consensus will be assessed using the same a priori threshold ($\geq 80\%$ agreement at scores ≥ 4).

Round 3 (conditional)

A third Delphi round will be conducted only if consensus has not been achieved on specific domains or methodological aspects after Round 2. In this round, panelists will vote on the final formulation of unresolved items. If consensus remains elusive after three rounds, the area of persistent disagreement will be transparently reported and discussed in the final manuscript, consistent with CREDES recommendations.¹⁷ The stability of responses across rounds will be assessed to determine whether additional iterations are likely to produce further convergence.²⁴

Definition of consensus

Consensus will be defined a priori as at least 80% of respondents rating an item with a score of 4 or 5 (agree or strongly agree) on the 5-point Likert scale. This threshold is consistent with widely adopted standards in Delphi research and has been recommended by multiple methodological reviews.^{15,21,25} Consensus will be further categorized as follows: 90-100% agreement = strong consensus; 80-89% agreement = consensus; 70-79% agreement = near-consensus (item retained for re-evaluation); $<70\%$ agreement = no consensus (item revised or removed). Items achieving strong consensus or consensus will be included in the final validated version of the SNR framework. Items achieving near-consensus will be revised and re-submitted in subsequent rounds. Items that fail to reach 70% agreement after the final round will be excluded from the core framework, although they may be retained as optional recommendations if supported by qualitative feedback.

Statistical analysis

Descriptive statistics will be used to summarize the results of each Delphi round. For each domain and each evaluation dimension (relevance, clarity, feasibility), the following metrics will be calculated: median Likert score, interquartile range (IQR), mean score with standard deviation and percentage agreement (proportion of ratings ≥ 4). Changes in agreement across rounds will be quantified and reported as absolute differences ($\Delta\%$).

Inter-rater reliability will be assessed using the intraclass correlation coefficient (ICC) with a two-way random effects model

for absolute agreement. ICC values will be interpreted according to established benchmarks: <0.50 = poor, 0.50 - 0.75 = moderate, 0.75 - 0.90 = good, and >0.90 = excellent agreement.²⁶ The stability of responses across Delphi rounds will be evaluated by comparing median scores and percentage agreement between consecutive rounds, with stability defined as a change of less than 15% in agreement for any given item.²⁴ Qualitative comments from participants will be analyzed using thematic synthesis to identify recurring themes, suggestions, and concerns. All quantitative analyses will be performed using R software, version 4.3.2 or later (R Foundation for Statistical Computing, Vienna, Austria; www.r-project.org).

Ethical considerations

The Delphi consensus process involves the voluntary participation of expert clinicians and researchers and does not involve patient data, biological samples, or clinical interventions. Participation will be anonymous during all Delphi rounds, and all responses will be handled confidentially. Nevertheless, the study protocol will be submitted for review to an institutional ethics committee to ensure compliance with international standards for research integrity, including the Declaration of Helsinki²⁷ and applicable data protection regulations (General Data Protection Regulation, GDPR, where applicable). Written informed consent will be obtained electronically from all participants prior to enrollment, with explicit information regarding data handling, anonymization procedures, and publication of aggregated results.

Data management and quality assurance

All survey data will be collected and stored on a secure, encrypted online platform with access restricted to authorized members of the Steering Committee. Data will be anonymized at the point of analysis to prevent identification of individual respondents. Regular data quality checks will be performed to ensure completeness and consistency of responses. The survey platform will include automated validation rules to minimize missing data and entry errors. A detailed audit trail of all methodological decisions, including revisions to survey items and rationale for changes between rounds, will be maintained and reported in the final manuscript.

Expected outcomes

The Delphi consensus process is expected to generate several key outputs. First, it will produce a consensus-endorsed version of the SNR framework, with core domains refined through iterative expert evaluation. Second, it will propose a standardized scoring system (0-24 points) with consensus-based quality thresholds to guide editorial assessment and peer review. Third, it will provide consensus-based recommendations for the adoption and integration of the SNR framework within journal editorial workflows. Additionally, the process will identify priorities for future research, particularly in relation to empirical validation, inter-rater reliability, and the framework's ability to discriminate between different levels of methodological quality in published narrative reviews. Finally, recommendations will be developed for discipline-specific adaptations and potential digital implementation of the SNR

assessment tool. Collectively, these outputs will lay the foundation for subsequent empirical validation studies and, potentially, broader dissemination initiatives. The extent to which these outputs may contribute to wider standardization of narrative review methodology will depend on the results of the consensus process and subsequent empirical testing.

Figure 1 illustrates the structured workflow of the modified Delphi consensus process, detailing the sequential rounds, decision nodes, and analytical steps leading to validation of the SNR framework.

Discussion

This protocol presents the methodological framework for an international modified Delphi consensus study designed to validate the SNR Framework. The study addresses a recognized gap in evidence synthesis methodology: the absence of a comprehensive, domain-based quality assessment tool specifically designed for

narrative reviews that balances methodological rigor with interpretative flexibility.^{3,5,28}

The Delphi technique was selected as the most appropriate methodology for this validation study for several reasons. First, narrative review methodology represents a field in which established empirical evidence is limited and expert judgment is essential for evaluating methodological constructs.^{11,12} Second, the iterative and anonymous nature of the Delphi process allows for convergence of expert opinion while minimizing the influence of dominant personalities, conformity pressure, and group dynamics that may affect face-to-face consensus methods.^{13,29} Finally, the electronic format facilitates the inclusion of geographically dispersed international experts, enhancing the representativeness and generalizability of the resulting consensus.^{14,17}

Several methodological tools currently exist for evaluating review quality, including SANRA for narrative reviews,⁵ PRISMA 2020 for systematic reviews,⁶ PRISMA-ScR for scoping reviews⁷ and AMSTAR 2 for systematic review quality assessment.³⁰ However, none of these tools provides a comprehensive, domain-

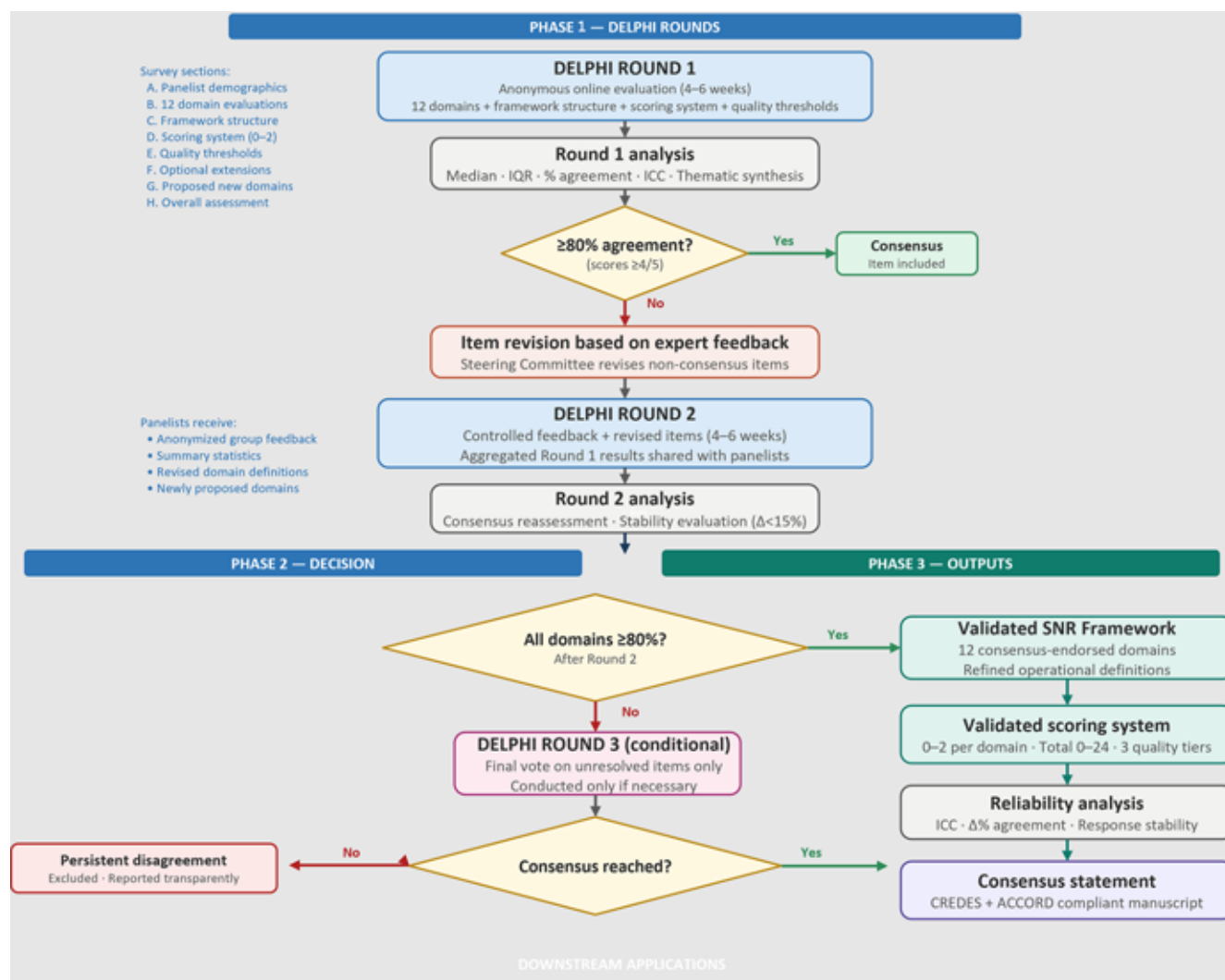


Figure 1. Workflow of the modified Delphi process for validation of the SNR framework, showing iterative consensus rounds, item revision and final outputs including framework validation, scoring system and consensus statement.

based framework specifically tailored to the unique methodological characteristics of narrative reviews. SANRA, while a pioneering contribution, comprises only six items and focuses primarily on reporting quality rather than methodological rigor. The SNR framework will extend beyond SANRA by incorporating elements of evidence stratification, mechanistic synthesis, quantitative awareness, and epistemological transparency, thereby addressing the full spectrum of methodological expectations for contemporary narrative reviews.

The Delphi validation process described in this protocol draws methodological inspiration from comparable initiatives in related fields. Notably, the modified Delphi approach has been successfully employed for the development of clinical assessment tools,³¹ diagnostic criteria³² and reporting guidelines, as exemplified by the ACCORD development process.¹⁸ The present study applies similar methodological rigor to the domain of narrative review methodology, an area that has received comparatively less attention in terms of formal consensus development.

The proposed study incorporates several methodological strengths. The use of a priori consensus thresholds and transparent decision rules enhances reproducibility and reduces the risk of post hoc adjustment of criteria.^{15,25} The multi-dimensional evaluation approach (relevance, clarity, feasibility) ensures that the framework is assessed not only for its conceptual soundness but also for its practical applicability. The inclusion of inter-rater reliability analysis using the ICC provides an additional layer of methodological rigor beyond simple percentage agreement.²⁶ Furthermore, the integration of qualitative thematic analysis alongside quantitative metrics enables a richer understanding of expert perspectives and facilitates the identification of areas requiring conceptual refinement.

Potential limitations

Several potential limitations should be acknowledged. First, the Delphi technique inherently relies on expert opinion and consensus does not necessarily equate to scientific truth.¹³ The consensus-endorsed framework will therefore require subsequent empirical validation, including inter-rater reliability studies using published narrative reviews and assessment of discriminative validity across different levels of methodological quality. The careful selection of experts, while necessary to ensure relevant expertise, may introduce selection bias. Efforts to mitigate this bias will include transparent reporting of selection criteria, demographic characteristics of the panel and response rates across rounds.¹⁷ Attrition between Delphi rounds may affect the stability and representativeness of the consensus. To address this, the study will predefine an acceptable minimum retention threshold of 70% of enrolled panelists for each round. Dropout rates will be reported transparently. Where feasible, available demographic and professional characteristics of non-completers will be compared with those of completers to assess the potential for attrition bias. Sensitivity analyses excluding items with high rates of missing data will be conducted and reported alongside the primary results.

The absence of empirical validation at the current stage represents an inherent constraint of the protocol design. The SNR framework has not yet been applied to published narrative reviews, and its real-world performance in terms of inter-rater agreement, discriminative sensitivity, and practical feasibility in editorial workflows remains to be established. Consensus among experts on the face validity and content adequacy of the framework does not substitute for evidence of its predictive validity or impact on review quality. These empirical questions will require dedicated post-consensus studies and should be explicitly addressed in future research.

The definition of consensus at 80% agreement, while widely adopted, remains somewhat arbitrary and different thresholds could yield different results.^{15,25} The sensitivity of the findings to alternative thresholds will be explored in supplementary analyses.

A further limitation that warrants explicit acknowledgment is the risk of over-standardization. Narrative reviews derive a substantial part of their scientific value from their interpretative flexibility, disciplinary sensitivity and capacity to synthesize heterogeneous evidence across contexts that resist uniform methodological treatment. A prescriptive scoring framework, if applied too rigidly, may inadvertently constrain this interpretative function, penalize legitimate methodological diversity, or impose a biomedical-style rigor that is inappropriate for humanities, social science, or emerging interdisciplinary fields. The SNR framework has been designed with this tension in mind, including optional rather than mandatory high-level extensions, but the risk of over-standardization should be monitored through qualitative feedback during the Delphi rounds and addressed explicitly in implementation guidelines.

Implications for practice and research

If validated, the SNR framework may offer significant benefits for multiple stakeholders. For authors, it provides clear methodological guidance for conducting and reporting high-quality narrative reviews. For editors and peer reviewers, it offers a structured appraisal tool analogous in function to CONSORT for randomized trials and PRISMA for systematic reviews, facilitating more consistent and objective evaluation processes. However, it should be explicitly acknowledged that Delphi-based consensus validation does not, in itself, confirm practical usability in real editorial or peer-review settings. The feasibility, inter-rater reliability, and discriminative validity of the SNR framework in applied contexts will require dedicated empirical studies, including field testing across diverse journals and disciplines, before widespread adoption can be recommended. For the broader scientific community, the adoption of the SNR framework may contribute to harmonizing narrative review standards, enhancing the credibility of narrative synthesis, and promoting its continued role in advancing translational and interdisciplinary research.²⁸

Future research should focus on empirical testing of the validated SNR framework across diverse disciplines, development of training resources and implementation toolkits, and evaluation of its impact on the quality of published narrative reviews through before-and-after studies or controlled comparisons.

Conclusions

This protocol describes a rigorous and transparent methodology for the international Delphi—validation of the SNR Framework. By adhering to established guidelines for Delphi study conduct and reporting, including CREDES and ACCORD, the proposed study aims to produce a consensus-endorsed methodological tool with the potential to advance narrative review quality beyond current standards. It is important to note that Delphi-based validation provides expert consensus, not empirical proof of effectiveness; the framework's real-world performance will require subsequent testing through inter-rater reliability studies and prospective application across disciplines. The successful completion of this consensus process will nonetheless provide a rigorous, transparent foundation for such empirical validation and for broader dissemination of the SNR framework.

References

- 1 Green BN, Johnson CD, Adams A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *J Chiropr Med* 2006;5:101-17.
- 2 Ferrari R. Writing narrative style literature reviews. *Med Writ* 2015;24:230-5.
- 3 Sukhera J. Narrative reviews: flexible, rigorous, and practical. *J Grad Med Educ* 2022;14:414-7.
- 4 Campbell M, Katikireddi SV, Sowden A, Thomson H. Lack of transparency in reporting narrative synthesis of quantitative data: a methodological assessment of systematic reviews. *J Clin Epidemiol* 2019;105:1-9.
- 5 Baethge C, Goldbeck-Wood S, Mertens S. SANRA-a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev* 2019;4:5.
- 6 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
- 7 Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018;169:467-73.
- 8 Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Social Res Methodol* 2005;8:19-32.
- 9 Popay J, Roberts HM, Sowden AJ, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A Product from the ESRC Methods Programme. 2006. Lancaster University.
- 10 Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 2009;26:91-108.
- 11 Humphrey-Murto S, Wood TJ, Gonsalves C, et al. The Delphi method. *Acad Med* 2020;95:168.
- 12 Barrett D, Heale R. What are Delphi studies? *Evid Based Nurs* 2020;23:68-9.
- 13 Niederberger M, Schifano J, Deckert S, et al. Delphi studies in social and health sciences-Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *PLoS One* 2024;19:e0304651.
- 14 Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: How to decide its appropriateness. *World J Methodol* 2021;11:116-29.
- 15 Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 2014;67:401-9.
- 16 Brady SR. Utilizing and adapting the Delphi method for use in qualitative research. *Int J Qual Methods* 2015;14.
- 17 Junger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and Reporting Delphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliat Med* 2017;31:684-06.
- 18 Gattrell WT, Logullo P, van Zuuren EJ, et al. ACCORD (ACcurate CONsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLoS Med* 2024;21:e1004326.
- 19 Simera I, Moher D, Hoey J, et al. A catalogue of reporting guidelines for health research. *Eur J Clin Invest* 2010;40:35-53.
- 20 ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals. Updated 2026. Available from: <https://www.icmje.org/recommendations/>
- 21 Manyara AM, Purvis A, Ciani O, et al. Sample size in multi-stakeholder Delphi surveys: at what minimum sample size do replicability of results stabilize? *J Clin Epidemiol* 2024;174:111485.
- 22 Crane D, Henderson EJ, Chadwick DR. Exploring the acceptability of a 'limited patient consent procedure' for a proposed blood-borne virus screening programme: a Delphi consensus building technique. *BMJ Open* 2017;7:e015373.
- 23 Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Med Res Methodol* 2005;5:37.
- 24 Holey EA, Feeley JL, Dixon J, Whittaker VJ. An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Med Res Methodol* 2007;7:52.
- 25 von der Gracht HA. Consensus measurement in Delphi studies. *Technol Forecast Soc Change* 2012;79:1525-36.
- 26 Koo TK, Li MY. A guideline of selecting and reporting intra-class correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-63.
- 27 World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191-4.
- 28 Snyder H. Literature review as a research methodology: An overview and guidelines. *J Bus Res* 2019;104:333-9.
- 29 Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval* 2007;12:10.
- 30 Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
- 31 Spranger J, Homberg A, Sonnberger M, Niederberger M. Reporting guidelines for Delphi techniques in health sciences: A methodological review. *Z Evid Fortbild Qual Gesundhwes* 2022;172:1-11.
- 32 Hohmann E, Beaufils P, Beiderbeck D, et al. Guidelines for designing and conducting Delphi consensus studies: an expert consensus Delphi study. *Arthroscopy* 2025;41:4208-24.e10.

Received: 1 April 2026; Accepted: 18 May 2026.

Contributions: all authors made a substantive intellectual contribution, read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflict of interest: the authors declare no competing interests relevant to the content of this article and all authors confirm accuracy.

Acknowledgments: the authors gratefully acknowledge the support of Fondazione Paolo Proccacci (Rome, Italy) for its contribution to the manuscript preparation and editing process, as well as for the valuable scientific discussions that supported the development of this work.

Publisher's note: all claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).