

Perspective

Open Access



Evidence evaluation in rare disease guidelines: a methodological perspective

Matt Bolz-Johnson^{1,2}, Thomas Kenny³, Charlotte Gaasterland⁴, Muhammad Imran Omar^{5,6}, Manon Engels⁷, Agnies van Eeghen^{8,9}, Iris den Uijl¹⁰, Willemijn Irvine¹¹

¹EURORDIS - Rare Diseases Europe, Brussels 1000, Belgium.

²Square Root Thinking GmbH, Cologne 50670, Germany.

³Patient Advocacy, CHIESI, Manchester M22 5LG, UK.

⁴Leiden University Medical Center, Leiden 2333 ZA, The Netherlands.

⁵Academic Urology Unit, Health Sciences Building, University of Aberdeen, Aberdeen AB24 3FX, UK.

⁶The Guideline Office, European Association of Urology, Arnhem 6803 AA, The Netherlands.

⁷Radboud University Medical Center, Nijmegen 6525 GA, The Netherlands.

⁸Kinderpolikliniek Klinische Genetica, Amsterdam University Medical Centre, Amsterdam 1105 AZ, The Netherlands.

⁹Heeren Loo Care Group, Amersfoort 3818 EN, The Netherlands.

¹⁰Erasmus MC Sophia Children's Hospital, Rotterdam 3015 GD, The Netherlands.

¹¹Qualicura, Breda 4811 CP, The Netherlands.

Correspondence to: Dr. Matt Bolz-Johnson, EURORDIS - Rare Diseases Europe, Brussels 1000, Belgium. E-mail: matt.johnson@eurordis.org

How to cite this article: Bolz-Johnson M, Kenny T, Gaasterland C, Omar MI, Engels M, van Eeghen A, den Uijl I, Irvine W. Evidence evaluation in rare disease guidelines: a methodological perspective. *Rare Dis Orphan Drugs J.* 2025;4:35. <https://dx.doi.org/10.20517/rdodj.2025.29>

Received: 13 May 2025 **First Decision:** 20 Jun 2025 **Revised:** 16 Jul 2025 **Accepted:** 22 Sep 2025 **Published:** 10 Dec 2025

Academic Editor: Daniel Scherman **Copy Editor:** Fangling Lan **Production Editor:** Fangling Lan

Abstract

This paper examines the methodological challenges of developing rare disease clinical guidelines and compares the standard Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach with an enhanced methodology tailored to rare disease constraints. Drawing on European Reference Network working groups, case studies, literature, and discussions from a EURORDIS webinar (April 2024), it identifies strategies to produce evidence-informed recommendations despite limited and heterogeneous data. The enhanced GRADE framework broadens search strategies, integrates qualitative synthesis, real-world evidence, and structured expert/patient input, and uses consensus methods such as Delphi processes and evidence-to-decision frameworks. This enables guideline developers to address sparse data, non-traditional research questions, and variable outcomes while maintaining transparency. For rare diseases, where conventional hierarchies of evidence are often unworkable, this adapted approach provides a flexible, pragmatic, and inclusive pathway. By leveraging



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



registries, expert consensus, and tailored evidence integration, it supports robust, context-sensitive guidelines that remain clinically relevant and improve care for underserved patients.

Keywords: Rare diseases, clinical guidelines, evidence synthesis, methodology

INTRODUCTION

The development of clinical practice guidelines (CPGs) seeks to identify and synthesize the highest quality evidence, typically through systematic reviews that form the basis for evaluating and formulating recommendations. For therapeutic interventions, evidence derived from randomized controlled trials (RCTs) is generally positioned at the apex of the evidence hierarchy, whereas case series and expert opinions are regarded as lower-quality sources^[1]. However, in the context of rare diseases - defined by a prevalence of fewer than 1 in 2,000 individuals^[2], collectively affecting approximately 30 million people in the European Union and an estimated 300 million globally^[3], the quantity of available evidence is frequently limited. Conducting RCTs for these conditions may often be impractical or ethically challenging^[4]. Small patient populations and a limited number of specialized centers impede the accumulation of robust, large-scale data. Moreover, potential participants and their caregivers may be overburdened, vulnerable, or otherwise unable to participate due to the severity and complexity of their conditions and associated disabilities. Consequently, these populations remain understudied. Given the persistent constraints on research pipelines, the likelihood of generating substantial additional evidence in the foreseeable future is low. Thus, for many rare diseases, the best available evidence frequently derives from clinical real-world experience documented in case series, structured expert consensus, or adaptive trial designs supported by real-world registry-derived data. In this manuscript, “real-world evidence” specifically denotes data from registries, patient databases, or observational studies. “Structured expert consensus” refers to formal methodologies (Delphi, modified Delphi, Evidence to Decision (EtD) scoring). “Unstructured clinical expertise” denotes experiential judgment outside formal processes. This distinction is applied consistently to avoid ambiguity.

Certain rare diseases are managed with targeted interventions, including orphan medicines specifically developed for such conditions. It is widely recognized that evaluating the evidence base for orphan medicinal products (OMPs) requires a tailored approach, distinct from that applied to interventions for more common conditions, due to the intrinsic challenges posed by small patient populations^[5]. Nevertheless, in the context of guideline development, the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) methodology remains the established standard for evaluating evidence quality^[6].

European Reference Networks (ERNs) are collaborative networks comprising centers of expertise and research units designed to consolidate knowledge and facilitate access to diagnosis, treatment, and highly specialized care for patients with rare and complex conditions^[5]. A core mandate of ERNs is to enhance care for patients with rare or complex diseases by developing CPGs and clinical decision support tools^[7]. The ERN experiences referenced here are not exhaustive, but focus on ERNICA, ERN GENTURIS, and ERN ITHACA, which have collectively produced guidelines spanning congenital anomalies, tumour risk syndromes, and genetic neurodevelopmental disorders. Methodologists working within these ERNs, such as ERNICA, ERN GENTURIS, and ERN ITHACA, have accrued substantial experience in applying the GRADE methodology under conditions characterized by limited and heterogeneous evidence. This practical experience has advanced the understanding of methodological challenges and informed the development of pragmatic solutions for generating high-quality guidelines in rare disease contexts.

Methodology 1: the standard GRADE methodology

The GRADE methodology is widely employed to assess the quality of evidence and guide the formulation of clinical recommendations through structured panel deliberations^[8,9]. This process constitutes the foundation of the EtD framework. Central to GRADE is the appraisal of evidence quality based on study design, whereby well-conducted RCTs are prioritized, and observational studies are typically downgraded due to potential biases or confounding^[8,9]. Evidence is evaluated on a per-outcome basis, incorporating dimensions such as risk of bias, inconsistency, indirectness, imprecision, and publication bias^[8]. The EtD framework subsequently integrates additional considerations, including patient values, resource utilization, and feasibility. However, in the context of rare diseases, the application of GRADE presents specific challenges owing to the scarcity of high-quality evidence^[10]. The limited availability of published studies - frequently small case series or observational reports - often results in uniformly low or very low evidence ratings, yielding weak or conditional recommendations^[10]. These ratings may not fully capture the value of structured expert consensus or unstructured clinical expertise.

For many rare and complex conditions, the total volume of peer-reviewed literature remains limited and is unlikely to expand substantially. Consequently, conventional evidence grading may not adequately support clinical guidance. This issue is particularly significant given that clinicians require evidence-based recommendations to inform critical care decisions, even in the absence of robust evidence. In practice, methodologists often encounter literature reviews that identify only a modest number of potentially relevant studies. Excluding these studies due to perceived methodological limitations may eliminate crucial information, thereby complicating the formulation of sound recommendations. Nevertheless, developing guidance remains imperative to avoid leaving rare disease communities without structured clinical direction.

The consistent reliance on lower levels of evidence, including expert opinion, is not merely unavoidable but essential in the context of rare diseases. For numerous conditions, RCTs may never be feasible, and well-documented case series may represent the best available evidence. Recognizing this, GRADE incorporates flexibility through its EtD framework and through adaptations such as the World Health Organization (WHO) and Confidence in the Evidence from Reviews of Qualitative Research (CERQual) approaches. This manuscript emphasizes the pragmatic application of these principles in rare disease contexts, underscoring the need for nuance and adaptability, particularly given the significant implications for patients and clinicians.

Methodologists participating in a EURORDIS-Rare Diseases Europe webinar held in April 2024 identified three principal challenges in synthesising evidence to support high-quality recommendations: (i) the limited quantity of available evidence; (ii) the heterogeneity of data; and (iii) issues related to the formulation of clinically relevant research questions. These themes are illustrated in [Figure 1](#).

Small amount of evidence and study types

The objective of evidence gathering for guideline development is typically to identify prospective studies. However, in practice, the available literature for rare diseases predominantly comprises retrospective observational studies. It is common to encounter publications titled “Disease “x”: a 20-year experience”, in which there is limited direct association between interventions and outcomes - precisely the type of information required for robust evidence-based guideline development. For example, in the ERNICA guideline on omphalocele, a rare abdominal wall defect, the systematic search identified approximately 700 potential studies, most of which were case reports, and fewer than ten were non-randomized trials^[11,12].

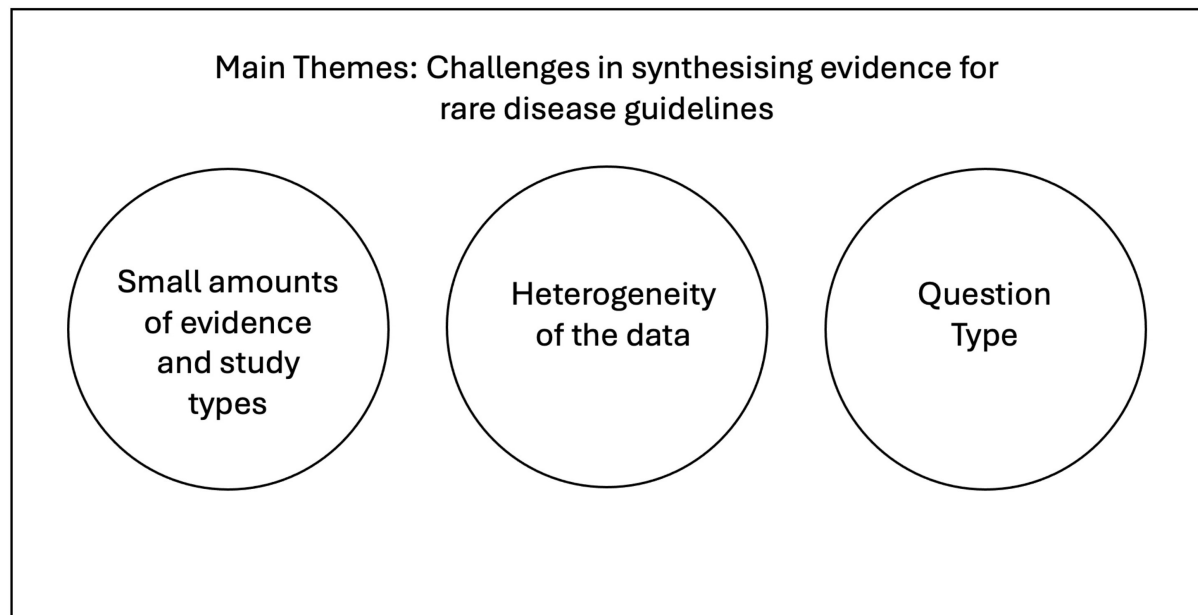


Figure 1. Main themes of the challenges in synthesizing evidence for rare disease guidelines.

Heterogeneity in data

Rare disease guidelines frequently contend with heterogeneity in outcomes, with diverse measures and scales employed to assess the same variables. This heterogeneity extends to differing time points for outcome assessment, such as evaluations conducted at 18 months, during school age, or at long-term follow-up, as well as variability in study populations, such as wide paediatric age ranges or mixed cohorts requiring the same surgical intervention^[11-13]. Given that GRADE evaluates certainty on a per-outcome basis, such heterogeneity complicates efforts to aggregate and synthesise the evidence.

Question type

Experts developing new rare disease guidelines often seek to address questions that do not conform to the Population, Intervention, Comparison, Outcome (PICO) framework, despite their clinical relevance. These questions may pertain to diagnostic pathways, care provision, or long-term surveillance strategies, rather than comparative analyses of interventions or prognostic factors. The challenge is further exacerbated by the insufficient evidence available to establish the evidence-based value of individual components within care processes.

The challenges in evaluating evidence for rare diseases are well recognized, prompting a fundamental question: Does the primary limitation lie in the scarcity of evidence, or is it attributable to the tools currently employed for grading evidence quality? Specifically, is GRADE sufficiently adaptable for low-prevalence patient populations, or does it require modifications to address the challenges outlined above? With more than 8,000 identified rare diseases, and numerous others remaining undiagnosed, an additional challenge is the constrained capacity of expert groups and clinical networks to develop guidelines that encompass the full breadth of rare diseases. ERNs are therefore compelled to prioritise guideline development. For instance, ERN ITHACA has established a prioritization framework that favours guidelines addressing cross-cutting needs such as neurocognitive, transitional, and behavioural comorbidities^[14]. Similarly, ERN GENTURIS has prioritized syndrome-specific guidelines^[15-19] and is expanding efforts to develop guidelines that address common aspects of genetic tumour risk syndromes.

The analysis drawn together in this manuscript is primarily on methodological experiences documented in ERNICA, ERN GENTURIS, and ERN ITHACA guidelines, as well as structured discussions held during the EURORDIS April 2024 webinar. Case study illustrations were derived from ERNICA gastroschisis guidelines^[11] ERN GENTURIS surveillance guidelines^[15,16], and ERN ITHACA prioritization frameworks^[14].

Methodology 2: adapted and enhanced approaches for rare diseases

GRADE is effective when ample high-level evidence is available. However, as noted, guideline developers frequently encounter challenges in rare disease contexts due to the limited quantity and heterogeneity of available evidence. In these situations, it remains essential to apply key principles of GRADE through an enhanced approach that accommodates very low-level evidence, often derived from small case series or case reports, thereby enabling structured synthesis on a per-outcome basis. The guideline panel subsequently assesses this evidence to determine whether to issue strong or conditional recommendations, followed by consensus-building processes such as Delphi methods. These panels also evaluate the balance of benefits and harms, integrating these considerations with the available evidence to establish the strength of recommendations. ERN guidelines have adopted this methodology^[11-13,15-19]. To mitigate unconscious bias and safeguard decision-making, ERN panels use anonymous scoring, multiple consensus rounds using DELPHI, and explicit declaration of conflicts of interest. Including patient representatives helps balance clinical perspectives.

An enhanced GRADE approach draws on all available evidence, published studies and real-world experience, both through structured expert consensus and registry-derived data, and employs broader, more inclusive search strategies to optimise the evidence base.

Limited evidence and study types

In rare disease research, it is common for authors to consolidate multiple stages of the diagnostic and therapeutic process within a single publication, unlike studies on more prevalent conditions, which typically focus on specific interventions. Consequently, the keywords utilized for literature searches in rare diseases need not be as narrowly defined as those employed for common diseases. This broader scope reflects the nature of rare disease guidelines, which often encompass an integrated continuum of diagnosis, care, and treatment.

When searches are conducted using individual PICO components or are narrowly focused on specific topics within rare disease guidelines, they frequently yield minimal or no evidence. To address this limitation, an emerging strategy involves conducting broader searches using only the disease term itself. This approach increases the number of retrieved articles and mitigates the risk of omitting pertinent evidence due to excessively restrictive search criteria.

Furthermore, in the context of rare disease guideline development, the role of the guideline panel is critical in supplementing published data with clinical expertise. It is increasingly recognized as essential to transition from unstructured expert opinion towards structured expert evidence, defined as systematically gathered insights through formal processes such as Delphi consensus exercises. This process includes soliciting information from medical experts and individuals with lived experience, which is particularly valuable when assessing the spectrum of anticipated benefits and harms associated with interventions.

Another complementary approach involves leveraging higher-quality evidence from more common diseases that share similar clinical characteristics or management challenges, effectively broadening the “P” in the PICO framework. For example, in the evaluation of congenital anomalies necessitating multiple surgical

procedures, where long-term outcomes specific to these anomalies may be under-researched, comparative data might be drawn from studies on the effects of intubation and anaesthesia in premature infants.

Evaluating the strength and quality of the evidence remains a cornerstone of developing high-quality guidelines, even under circumstances of evidence scarcity. In the context of rare diseases, it is critical to assess not only the available evidence but also the plausibility of obtaining higher-quality evidence within a reasonable timeframe. Clinicians and patients can often reasonably judge that, for example, if the best available evidence consists of a case series of 40 patients, it is improbable that substantially larger studies will be conducted in the near future.

It is equally important to recognize that guidelines are not static instruments; they require periodic review, updating, and re-contextualization to ensure continued relevance as new evidence emerges. This iterative process underpins the long-term quality and applicability of clinical recommendations.

Heterogeneity in data

To facilitate effective data synthesis, it is essential to establish a clear working definition of key search terms at the outset of the process. Variability in clinical definitions is common among practitioners; therefore, adopting and consistently applying a unified definition enhances methodological rigour throughout the systematic review process.

A major challenge in rare disease research is the inconsistent reporting of outcomes across studies, resulting in a lack of standardization that hampers meaningful comparisons and synthesis. Addressing this issue may necessitate a more holistic approach that considers a broader array of outcomes, rather than focusing narrowly on individual effect estimates. This strategy enables a more comprehensive understanding of the collective evidence, encompassing the full range of potential benefits and harms associated with interventions in rare diseases. Looking forward, the adoption of Core Outcome Sets in future clinical trials could mitigate this challenge by standardising outcomes^[20].

In rare disease guideline development, formal quantitative meta-analyses are uncommon. Instead, qualitative analyses are typically employed, often involving multiple parallel assessments of the same studies. Guideline methodologists assign evidence ratings, which are then reviewed by the guideline panel. This process relies on the panel's clinical expertise and judgment to formulate final recommendations. The GRADE EtD framework provides a valuable tool for structuring these deliberations.

To support EtD discussions, panels review summarized evidence and assess various outcome measures using structured surveys or expert deliberations^[9]. Consensus is generally achieved through modified Delphi processes involving multiple survey rounds. Additionally, in-person or virtual workshops may be convened to facilitate discussion. Patient experts can also contribute valuable data through focus groups or interviews, ensuring that these perspectives are incorporated into the evidence-to-decision deliberations.

Question type

It is critical for rare disease guidelines to draw upon all available evidence to inform recommendations. This becomes particularly relevant for questions concerning follow-up protocols, diagnostic pathways, or specific testing strategies. Such questions may not align with traditional comparative frameworks but can be addressed by careful literature review designs that include observational studies, thereby supporting the formulation of recommendations through structured Delphi consensus methods.

Conducting broader literature searches enriches the evidence base for the EtD process, facilitating an enhanced approach that integrates real-world evidence drawn from both registry-derived data and structured expert consensus, as well as unstructured clinical expertise, and patient perspectives^[21].

Under certain circumstances, the GRADE EtD framework explicitly permits the formulation of strong recommendations even when based on low-quality evidence, provided additional supporting factors are transparently documented. A guideline panel may thus issue a strong recommendation grounded in comprehensive judgment and justification, despite the formal evidence rating constraints. In practice, thresholds for issuing recommendations in the context of low-certainty evidence rely on a structured balance-of-reasoning approach. Panels assess: (i) the magnitude and consistency of observed clinical effect, even from small or uncontrolled studies; (ii) corroboration from structured expert consensus processes (e.g., Delphi); and (iii) alignment with patient-reported outcomes and preferences. Evidence is then graded according to the following criteria:

- Strong - Expert consensus and consistent evidence across sources (e.g., case series plus registry data).
- Moderate - Consensus is reached but evidence is mixed, evolving, or heterogeneous.
- Weak - Recommendations based on majority expert opinion without consistent supporting evidence.

By transparently documenting the rationale in the EtD tables, guideline developers ensure that even strong recommendations issued under low-quality evidence are justified, reproducible, and appropriately qualified.

Real world experience and data

For each intervention under consideration, it is essential to examine critical care outcomes alongside the benefit-to-harm ratio. In situations where empirical evidence on pivotal outcomes is lacking, guideline panels should draw upon their collective clinical experience to inform these assessments^[21]. Panels must also account for the values and preferences of the patient population^[22], and systematically discuss the strength of emerging recommendations^[23]. This process is enriched by qualitative patient insights gathered through structured methods, which can substantively guide the panel's deliberations^[21,23]. Identifying gaps in the literature within the guideline text is also crucial to inform future research agendas. Where evidence is conflicting or insufficient, it is imperative to employ formal consensus methodologies, such as Delphi or modified Delphi processes, to achieve agreement among panel members.

In rare disease guideline development, incorporating real-world evidence and structured expert consensus is indispensable. Data derived from patient registries or disease-specific databases should be regarded as critical supplementary evidence, particularly when published studies do not sufficiently address all pertinent clinical questions^[24].

ERNs employ a range of structured mechanisms to operationalize an enhanced application of the GRADE framework tailored to the specific challenges of rare diseases. These include the formation of dedicated multidisciplinary working groups comprising clinical experts, methodological specialists, and patient representatives, tasked with systematically evaluating the available evidence and drafting guideline content. Consensus is achieved through formal processes such as modified Delphi methodologies, structured surveys, and iterative panel discussions designed to ensure methodological rigour and transparency in the incorporation of expert judgment. Additionally, ERNs implement scheduled protocols for the periodic review and updating of guidelines, facilitating the integration of new evidence and the continuous

Table 1. Comparison: standard GRADE vs. enhanced GRADE approaches for rare diseases

Aspect	Standard GRADE approach	Enhanced GRADE approach for rare diseases
Purpose	Evaluate high-certainty evidence, typically RCT-based, to inform strong or conditional recommendations	Accommodate low-volume, heterogeneous, and real-world registry-derived data where RCTs are infeasible
Evidence Base	Prioritises systematic reviews of RCTs; observational studies generally downgraded	Incorporates case series, observational studies, registry data, and structured clinical expertise
Search Strategy	Narrow, PICO-focused searches tailored to specific interventions and outcomes	Broader disease-based searches to capture dispersed or limited evidence
Real-World Evidence	Considered supplementary; typically assigned lower confidence	Integrated as a primary evidence source, both registry-derived data and structured expert consensus, especially when published studies are scarce
Expert Input	Used to interpret data; secondary to formal evidence	Treated as structured “expert evidence” systematically elicited via panels or Delphi processes
Patient Preferences & Values	Included in EtD considerations, often via existing qualitative studies	Directly elicited through structured methods (e.g., focus groups, interviews)
Handling of Heterogeneity	Seeks homogeneity; heterogeneity generally lowers certainty	Accepts heterogeneity; focuses on identifying core outcomes and qualitative synthesis
Outcome Assessment	Prefers quantitative meta-analysis with outcome-specific grading	Employs qualitative synthesis; encourages Core Outcome Sets to improve future consistency

Notes: adapted from summary of the discussion between ERN methodological expertise, EURORDIS webinar (April 2024).

alignment of recommendations with current clinical practice. Collectively, these strategies enable ERNs to adapt GRADE principles effectively, combining systematic evidence appraisal with registry-derived data and structured expert consensus to develop robust, context-appropriate guidance for rare disease care.

Global relevance

Although this methodological perspective is primarily informed by experiences within ERNs, the challenges it addresses - such as limited patient populations, data heterogeneity, and the necessity of structured expert consensus - are equally pertinent in low- and middle-income countries (LMICs). Health systems in LMICs often contend with similar limitations, including constrained research infrastructures and scarce trial data. Thus, adapting an enhanced application of GRADE that systematically integrates real-world evidence and formalized expert input could facilitate robust, context-sensitive guideline development beyond Europe, ultimately improving care for underserved rare disease populations globally.

The direct comparison between the standard GRADE approach and an enhanced GRADE approach that is relevant to evaluating the evidence for rare disease guidelines can be found in [Table 1](#) below.

DISCUSSION

The GRADE EtD framework represents a robust and extensively validated system for formulating clinical recommendations, particularly in contexts where large volumes of high-quality evidence are available. However, additional methodological exploration is warranted for situations involving limited, heterogeneous evidence derived from non-randomized designs that cannot readily be synthesized across outcomes. There remains a critical clinical imperative to develop recommendations that support practitioners and address the substantial unmet medical needs inherent to rare diseases.

Defining and incorporating diverse sources of evidence is essential for developing rigorous rare disease guidelines. While such guidelines should systematically draw upon published evidence, they must also explicitly integrate registry-derived data and structured expert consensus as a valid form of evidence. Guideline development in rare and complex conditions often necessitates assembling all available data -

published studies, observational reports, expert experience, and patient insights - to ensure comprehensive decision-making.

The GRADE system is inherently designed to accommodate more than published data alone; it explicitly combines evidence with expert judgments from both clinicians and patient representatives within its EtD framework^[9]. In the context of rare diseases, it is imperative to involve these stakeholders through structured and formalized processes to mitigate potential biases. Employing rigorous methodologies that critically evaluate both published data and experiential knowledge - such as hybrid GRADE-and-Delphi approaches - can strengthen the robustness of recommendations. While Delphi or other consensus-building techniques can help minimise bias, there remains an unresolved question regarding the optimal balance between formal evidence and the substantial insights offered by expert clinical experience. Indeed, for rare diseases, this real-world expertise may, in specific contexts, be of even greater practical value than the limited published evidence. Developing systematic approaches for integrating this form of knowledge is a continuing methodological priority.

Developers of rare disease guidelines increasingly seek practical guidance on these issues, including criteria for determining when a formal guideline is warranted *vs.* when a consensus statement may be more appropriate. There is also a need for clear methodologies to incorporate and appraise registry data relative to data derived from RCTs, as well as to determine the evidentiary contribution of single case studies. Addressing these challenges is crucial to advancing the quality and applicability of future guidelines for rare and complex conditions.

“Real world evidence” is progressively recognized as an indispensable element for strengthening clinical guidelines. Much of the routine care provided in specialized centers, particularly for rare diseases, remains unpublished, yet constitutes a vital component of clinical practice. Integrating these insights into guideline development is essential to complement the published evidence base. The effectiveness of such care should be documented and leveraged through the experiential knowledge contributed by expert panels.

Emerging frameworks such as real world evidence studies (STaRT-RWE), and International Society for Pharmacoeconomics and Outcomes Research (ISPOR)/International Society for Pharmacoepidemiology (ISPE) principles offer structured approaches to appraise real-world evidence and patient registry data, which can complement rare disease guideline development.

Translating data and clinical experience into actionable recommendations requires expert judgment, particularly since available evidence may not align neatly with conventional PICO formulations. This is especially true for rare diseases, where evidence often does not conform to standard hierarchies and relies more heavily on clinical consensus. In many cases, registry-derived data derived from clinical experience and expert opinion may provide a more compelling and contextually appropriate foundation for recommendations than the limited published literature.

In light of these considerations, it may be argued that the traditional pyramid of evidence quality^[1,25] should be conceptually inverted for rare diseases. Under such a framework, the highest level of evidence could derive from structured expert consensus within ERNs, reflecting both the heterogeneous nature of rare conditions and the depth of clinical data. The scarcity of large-scale studies, due to small-affected populations and a limited number of centers capable of generating robust data, necessitates a recalibration of how evidence is valued in these contexts.

A critical aspect of incorporating real-world evidence into guidelines and consensus processes involves utilizing data from patient registries. Ensuring the reliability of such data requires rigorous validation procedures, as the credibility and precision of registry data are fundamental to supporting clinical decision-making. While rare disease registries remain an under-utilized resource, they possess significant potential to inform future guideline development.

For more common conditions, the methodological limitations of RCTs, particularly double-blind trials, are well-documented. In the context of rare diseases, it is even more imperative to pursue the highest feasible methodological quality and to rigorously limit bias within constrained study designs. For example, when the most viable study design is a single-arm trial with a small sample size, it is incumbent upon researchers to adopt rigorous design and analysis strategies to maximize the reliability of findings.

In addition to methodological challenges associated with evidence generation, rare disease care is frequently complicated by issues of access and affordability of OMPs. Despite advances in the development of targeted therapies, substantial barriers to equitable access persist even within high-income countries. Rare disease drugs are often associated with considerable financial burden for both healthcare systems and patients, potentially limiting their availability and uptake in clinical practice^[26]. Beyond methodological challenges, the financial realities of implementing rare disease guidelines cannot be overlooked. Access to promising but high-cost therapies poses a persistent barrier, and these economic pressures risk limiting not only the uptake of guideline recommendations but also the collection of much-needed real-world data. This underscores the importance of developing guideline methodologies that not only accommodate the inherent evidence limitations in rare diseases but also explicitly consider economic and implementation factors. Incorporating assessments of cost-effectiveness and resource allocation into the guideline process may therefore be essential to ensure that recommendations are both clinically sound and practically feasible across diverse healthcare settings. Innovative funding mechanisms and conditional reimbursement schemes - linking coverage to ongoing evidence generation - have been proposed as ways to balance timely access with the need for robust effectiveness data^[27].

Evidence for guidelines vs. HTA decisions

An important question pertains to whether evidence grading standards in health technology assessments (HTAs) should align with those employed in clinical guidelines, especially regarding the relative weighting of RCTs *vs.* other types of evidence in rare disease contexts. It is essential to delineate these distinct objectives: HTAs primarily focus on reimbursement decisions based on the best available evidence, whereas clinical guidelines aim to support clinical decision-making, often necessitating more nuanced and context-sensitive considerations.

Unlike CPGs, which integrate registry-derived data to guide bedside decisions, HTA processes (such as under the EU JCA regulation) focus on reimbursement decisions that often require different standards, emphasizing external validity and comparative efficacy. It is critical to recognize that while clinical guidelines seek to optimize individual patient care by incorporating diverse evidence sources and expert judgment, HTAs primarily assess comparative effectiveness and cost considerations to inform reimbursement decisions. This divergence necessitates distinct methodological approaches, particularly in rare diseases, where guideline development may rely more heavily on real-world data and structured expert consensus than would typically be acceptable within HTA frameworks.

Adaptive trial designs and the incorporation of real-world data as complementary data sources - particularly where RCTs are infeasible - are increasingly recognized by regulatory authorities^[24,28,29]. In the case of HTAs

for OMPs, recent Joint Clinical Assessments (JCAs) underscore the requirement for the highest-quality research to establish safety and efficacy^[30]. However, they also acknowledge the validity of single-arm studies in specific scenarios, such as ultra-rare conditions or settings with ethical constraints^[29]. This evolution highlights the growing adaptability of HTAs, which, depending on the rarity of the condition and the available data, may accept lower levels of evidence and integrate real-world data.

For guideline development utilizing the GRADE framework, the system necessitates a more context-dependent appraisal of evidence. Conventionally, single-arm studies are assigned lower grades compared to randomized, double-blind trials, especially when evaluating objective outcomes. However, the context and nature of the outcomes assessed are pivotal. For instance, in studies of fluctuating symptoms, single-arm designs may be deemed low-quality evidence. Conversely, in cases where patients with severe conditions demonstrate rapid and substantial improvements following an intervention, even single-arm studies may constitute compelling evidence. This variability underscores the need for a flexible, context-driven approach when grading evidence strength, particularly within rare disease contexts where conventional evidence hierarchies may be insufficient.

The GRADE methodology itself has evolved through iterative modifications, guided by the GRADE Working Group in collaboration with entities such as the WHO, the Guidelines International Network (GIN), and national bodies including National Institute for Health and Care Excellence (NICE) and the Cochrane Collaboration (Cochrane). There is now a pressing imperative to further adapt GRADE to meet the unique methodological demands associated with rare diseases, as elaborated in this manuscript.

CONCLUSION

The GRADE methodology serves not only as a framework for evaluating published evidence but also explicitly incorporates expert judgment, acknowledging the essential role of clinical expertise in guideline development^[9]. When appropriately applied, this approach enables experts to bridge gaps between available data and clinical recommendations, ensuring that resulting guidelines are transparent, well-substantiated, and informed by a holistic appraisal that includes registry-derived data and structured expert consensus.

The existing GRADE EtD framework integrates expert experience alongside formal evidence^[9]. However, this integration can be further strengthened through structured consensus-building methodologies, which provide additional safeguards against bias and facilitate balanced incorporation of evidence-based practices and expert insights. In the context of rare diseases, the combined weight of published data and expert experience assumes particular importance. When systematically assessed and integrated, this combined evidence base optimises the development of guidelines that reflect the full spectrum of scientific and clinical knowledge. By formally embedding expert experience alongside empirical evidence, guidelines can more effectively support healthcare professionals in delivering optimal care tailored to the complex needs inherent to rare disease populations.

These insights may inform similar methodological adaptations in LMICs, where data constraints often parallel those observed in European rare disease contexts. Evaluating evidence for rare disease guidelines must remain grounded in comprehensive systematic reviews of all relevant data. The principal challenges in synthesising high-quality recommendations for rare diseases stem from limited evidence volume, data heterogeneity, including variations in outcomes, time points, and study populations, and the nature of clinically pertinent research questions.

This paper has summarized practical strategies for mitigating the challenges associated with evaluating evidence for rare diseases, such as:

1. Conduct a broader search using only the disease term instead of specific PICO questions;
2. Utilize the real-world evidence from registry-derived data and structured expert consensus, as well as comparative evidence from other conditions;
3. Define broad clinical outcome definitions early in the process and adopt a more holistic approach to evaluating the results; and
4. Complete not only an assessment of the available evidence but also consider the likelihood of "stronger quality" evidence becoming available within a reasonable timeframe.

Taking an “enhanced GRADE” approach mitigates the challenges associated with rare disease research and ensures that guidelines are both evidence-based and clinically relevant to the complex needs of rare disease populations.

DECLARATIONS

Acknowledgments

Thank you to European Reference Networks & the EAU Guideline Office for their support and contribution to the webinar.

Authors' contributions

Design, data analysis, writing: Bolz-Johnson M

Manuscript editing, manuscript revision: Bolz-Johnson M, Kenny T, Gaasterland C, Omar MI, Engels M, van Eeghen A, den Uijl I, Irvine W

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Bolz-Johnson M is the Editorial Board Member of *Rare Disease and Orphan Drugs Journal*. Matt Bolz-Johnson was not involved in any steps of editorial processing, notably including reviewers' selection, manuscript handling and decision making, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg.* 2011;128:305-10. DOI PubMed PMC
2. European Commission. Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products; 2000. Available from: <https://eur-lex.europa.eu/eli/reg/2000/141/oj> [Last accessed on 20 Oct 2025]
3. Nguengang Wakap S, Lambert DM, Olry A, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet.* 2020;28:165-73. DOI PubMed PMC
4. Park Y, Fullerton HJ, Elm JJ. A pragmatic, adaptive clinical trial design for a rare disease: The FOcal Cerebral Arteriopathy Steroid (FOCAS) trial. *Contemp Clin Trials.* 2019;86:105852. DOI PubMed PMC
5. European Commission. Evaluation of the medicines for rare diseases and children legislation. 2020. Available from: https://health.ec.europa.eu/medicinal-products/medicines-children/evaluation-medicines-rare-diseases-and-children-legislation_en#documents [Last accessed on 20 Oct 2025]
6. Schünemann HBJ, Guyatt G, Oxman A. GRADE handbook. The GRADE Working Group; 2013. Available from: <https://gdt.gradepro.org/app/handbook/handbook.html> [Last accessed on 20 Oct 2025]
7. Tumiene B, Graessner H. Rare disease care pathways in the EU: from odysseys and labyrinths towards highways. *J Community Genet.* 2021;12:231-9. DOI PubMed PMC
8. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol.* 2011;64:383-94. DOI
9. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol.* 2013;66:726-35. DOI
10. European Reference Networks. Handbook #4: Methodology for the development of clinical practice guidelines for rare or low-prevalence and complex diseases; 2020. Available from: https://health.ec.europa.eu/publications/european-reference-network-clinical-practice-guidelines-and-clinical-decision-support-tools_en [Last accessed on 20 Oct 2025]
11. Burgos CM, Irvine W, Vivanti A, et al. European reference network for rare inherited congenital anomalies (ERNICA) evidence based guideline on the management of gastroschisis. *Orphanet J Rare Dis.* 2024;19:60. DOI PubMed PMC
12. Neville JJ, den Uijl I, Irvine W, Eaton S, Gottrand F, Hall NJ. Development of a core outcome set for paediatric achalasia: a joint ERNICA, ESPGHAN and EUPSA study protocol. *BMJ Paediatr Open.* 2025;9:e003130. DOI PubMed PMC
13. Hulscher J, Irvine W, Conforti A, et al. European reference network for inherited and congenital anomalies evidence-based guideline on surgical aspects of necrotizing enterocolitis in premature neonates. *Neonatology.* 2025;122:376-84. DOI PubMed PMC
14. Haneveld MJ, Oerbekke MS, Szakszon K, Cornel MC, Gaasterland CMW, Van Eeghen AM. Priority-setting criteria for clinical practice guideline development on rare genetic neurodevelopmental disorders: a Delphi study within the European Reference Network ITHACA. *J Clin Epidemiol.* 2025;182:111761. DOI PubMed
15. Carton C, Evans DG, Blanco I, et al. ERN GENTURIS tumour surveillance guidelines for individuals with neurofibromatosis type 1. *EClinicalMedicine.* 2023;56:101818. DOI PubMed PMC
16. Evans DG, Mostaccioli S, Pang D, et al. ERN GENTURIS clinical practice guidelines for the diagnosis, treatment, management and surveillance of people with schwannomatosis. *Eur J Hum Genet.* 2022;30:812-7. DOI PubMed PMC
17. Frebourg T, Bajalica Lagercrantz S, Oliveira C, Magenheimer R, Evans DG; European Reference Network GENTURIS. Guidelines for the Li-fraumeni and heritable TP53-related cancer syndromes. *Eur J Hum Genet.* 2020;28:1379-86. DOI PubMed PMC
18. Geilswijk M, Genuardi M, Woodward ER, et al. ERN GENTURIS clinical practice guidelines for the diagnosis, surveillance and management of people with Birt-Hogg-Dubé syndrome. *Eur J Hum Genet.* 2024;32:1542-50. DOI PubMed PMC
19. Tischkowitz M, Colas C, Pouwels S, Hoogerbrugge N; PHTS Guideline Development Group; European Reference Network GENTURIS. Cancer surveillance guideline for individuals with PTEN hamartoma tumour syndrome. *Eur J Hum Genet.* 2020;28:1387-93. DOI PubMed PMC
20. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012;13:132. DOI PubMed PMC
21. Jandhyala R. The multiple stakeholder approach to real-world evidence (RWE) generation: observing multidisciplinary expert consensus on quality indicators of rare disease patient registries (RDRs). *Curr Med Res Opin.* 2021;37:1249-57. DOI PubMed
22. Murad MH, Montori VM, Guyatt GH. Incorporating patient preferences in evidence-based medicine. *JAMA.* 2008;300:2483-4. DOI PubMed
23. Tringale M, Stephen G, Boylan AM, Heneghan C. Integrating patient values and preferences in healthcare: a systematic review of qualitative evidence. *BMJ Open.* 2022;12:e067268. DOI PubMed PMC
24. Liu J, Barrett JS, Leonardi ET, et al. Natural history and real-world data in rare diseases: applications, limitations, and future perspectives. *J Clin Pharmacol.* 2022;62 Suppl 2:S38-55. DOI PubMed PMC
25. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS Med.* 2008;5:e67. DOI PubMed PMC

26. Chaudhary A, Kumar V. Rare diseases: a comprehensive literature review and future directions. *J Rare Dis.* 2025;4:99. DOI
27. Ng QX, Ong C, Chan KE, et al. Comparative policy analysis of national rare disease funding policies in Australia, Singapore, South Korea, the United Kingdom and the United States: a scoping review. *Health Econ Rev.* 2024;14:42. DOI PubMed PMC
28. Chen J, Gruber S, Lee H, et al. Use of real-world data and real-world evidence in rare disease drug development: a statistical perspective. *Clin Pharmacol Ther.* 2025;117:946-60. DOI
29. European Medicines Agency. Reflection paper on establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing authorisation application. 2024. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-establishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketing-authorisation-application_en.pdf [Last accessed on 20 Oct 2025]
30. Official Journal of the European Union. Regulation (EU) 2021/2282 of the European parliament and of the council; 2021. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021R2282> [Last accessed on 20 Oct 2025]