



OPEN ACCESS

EDITED BY

Filippo Drago,
University of Catania, Italy

REVIEWED BY

Luis Laranjeira,
Ordem dos Médicos, Portugal
Lucia Gozzo,
University of Catania, Italy

*CORRESPONDENCE

Lotte Delemarre,
✉ lotte.delemarre@kuleuven.be

RECEIVED 18 November 2025

REVISED 04 January 2026

ACCEPTED 09 January 2026

PUBLISHED 12 March 2026

CITATION

Delemarre L, Huys I, Van Dyck W and Simoens S (2026) Addressing clinical uncertainties in ATMP reimbursement: a review of methodological guidelines and European practice. *Front. Pharmacol.* 17:1749386. doi: 10.3389/fphar.2026.1749386

COPYRIGHT

© 2026 Delemarre, Huys, Van Dyck and Simoens. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Addressing clinical uncertainties in ATMP reimbursement: a review of methodological guidelines and European practice

Lotte Delemarre^{1*}, Isabelle Huys¹, Walter Van Dyck² and Steven Simoens¹

¹KU Leuven Department of Pharmaceutical and Pharmacological Sciences, Leuven, Belgium, ²Healthcare Management Centre, Vlerick Business School, Ghent, Belgium

Introduction: Advanced Therapy Medicinal Products (ATMPs) often present substantial clinical uncertainties at the time of reimbursement evaluation, particularly due to the lack of appropriate comparators and the absence of long-term clinical endpoints. This study primarily examined two methodological areas relevant to these challenges: indirect treatment comparisons (ITCs) and surrogate endpoints. In addition, the review was supplemented with an assessment of innovative trial designs to explore how emerging approaches may contribute to evidence generation for ATMPs.

Methods: A structured literature review was conducted using PubMed and Embase, combining keywords such as “Health Technology Assessment,” “ATMP,” “Indirect Treatment Comparison,” and “Surrogate Endpoint,” supplemented with grey literature searches. European pharmacoeconomic guidelines were analyzed using the ISPOR database and referenced national documents.

Results: By April 2025, 27 ATMPs had received European Medicines Agency approval; 20 relied on single-arm trials, and 19 used surrogate endpoints. Single-arm trials limit direct comparative effectiveness assessment, requiring alternative approaches such as ITCs. Conventional ITCs need a common comparator, which is unavailable in single-arm evidence, while population-adjusted indirect comparisons (PAICs) offer a potential solution but depend on strong assumptions and are sensitive to unmeasured confounding. Guidelines generally accept the use of ITCs when direct comparative evidence is unavailable. However, they provide limited guidance on the separate question of when RCTs should be considered infeasible. Surrogate endpoints are widely used due to curative aims of ATMPs, but their validation typically requires RCT evidence and is highly context-specific. Guideline acceptance criteria for surrogate endpoints vary. Some require robust validation demonstrating a causal relationship, while others allow correlation-based or context-specific justification, particularly when direct evidence is limited or the technology addresses rare or serious conditions. Innovative trial designs may help generate more robust evidence in these challenging contexts.

Discussion: The study highlights the need to refine methods for handling single-arm trials, projecting long-term outcomes, and using surrogate endpoints in

ATMP evaluations. While some clinical uncertainty is inevitable, healthcare systems must manage it effectively to ensure timely, equitable access without creating additional barriers.

KEYWORDS

advanced therapy medicinal products, health technology assessment, indirect treatment comparison, reimbursement, surrogate endpoint

1 Introduction

Advanced Therapy Medicinal Products (ATMPs), as defined by the European Medicines Agency (EMA), are medicinal products based on genes, tissues, or cells (European Parliament and Council of the European Union, 2007). This category encompasses gene therapies, somatic-cell therapies, tissue-engineered medicines, and advanced therapies combined with medical devices. As of April 2025, 27 ATMPs had received marketing authorisation within the European Union (European Medicines Agency, 2025).

ATMPs have prompted a significant shift in pharmaceutical development by introducing treatments that often aim to cure (Detela and Lodge, 2019). Conventional medicines are typically designed to alleviate symptoms and slow further progression, often requiring continuous administration over a patient's lifetime. In contrast, many ATMPs can achieve lasting therapeutic effects with a single dose (Vestergaard et al., 2013). They are also frequently tailored to individual patients, moving beyond the standardized "one-size-fits-many" model of traditional drug development (Ledford, 2025).

These distinctive features present substantial challenges for the clinical evaluation of ATMPs. A prominent challenge is the frequent absence of a relevant comparator (Pinho-Gomes and Cairns, 2022). ATMPs are often developed for life-threatening diseases currently lacking effective treatments, with only best supportive care available. For these conditions, where no other therapeutic options exist, it can be considered unethical to withhold a potentially life-saving ATMP from patients by assigning them to a placebo or an inferior comparator arm in a clinical trial (Cargill et al., 2021). Furthermore, many ATMPs hold curative potential, meaning their full clinical benefit may only manifest years or even decades post-treatment. These challenges are further complicated by the fact that ATMPs often target rare diseases with very small patient populations (Hanna et al., 2016). This inherent rarity severely limits the feasibility of conducting robust comparative studies, thus contributing to considerable clinical uncertainties.

These challenges in the clinical evaluation pose considerable difficulties for subsequent reimbursement decisions (Iglesias-Lopez et al., 2021a). Health Technology Assessment (HTA) bodies typically assess the added therapeutic value of new interventions by comparing them to existing standards of care. This evaluation predominantly relies on comparative effectiveness data, often derived from RCTs (Fontrier et al., 2022). The absence of direct comparators for many ATMPs, coupled with the reliance on small, often single-arm studies, complicates the quantification of the incremental health benefit in a manner consistent with established HTA methodologies (Gozzo et al., 2021). Another key challenge is the frequent lack of long-term data on the durability of clinical outcomes. This uncertainty limits the ability to fully assess their long-term value within the typical timeframes used in HTA. A possible solution for this is the use of surrogate endpoints, which can provide earlier indications of a therapy's potential long-term effectiveness (Gladwell et al., 2024). While these therapies often entail

high upfront costs, their potential long-term health benefits and cost offsets remain unproven at the time of evaluation.

To our knowledge, two articles to date have provided an overview of the clinical uncertainties associated with ATMPs alongside potential methodological solutions (Iglesias-Lopez et al., 2021b; Qiu et al., 2022). In this article, we aim to delve deeper into the methodological strategies that can address these uncertainties and explore how various national pharmacoeconomic guidelines within Europe have responded to them. We specifically focus on two major challenges: the frequent absence of a relevant comparator and the lack of long-term clinical endpoints at the time of evaluation. These topics are especially important as an increasing number of single-arm trials, ITCs, and studies using surrogate endpoints are being submitted to reimbursement agencies (Patel et al., 2021; Harricharan et al., 2021; Grigore et al., 2020; Wheaton and Bujkiewicz, 2025). Methodologies for managing small patient populations, while relevant, were not a primary focus of this analysis, given the recent extensive literature on the topic (Day et al., 2018; Friede et al., 2018).

2 Methods

2.1 Review design

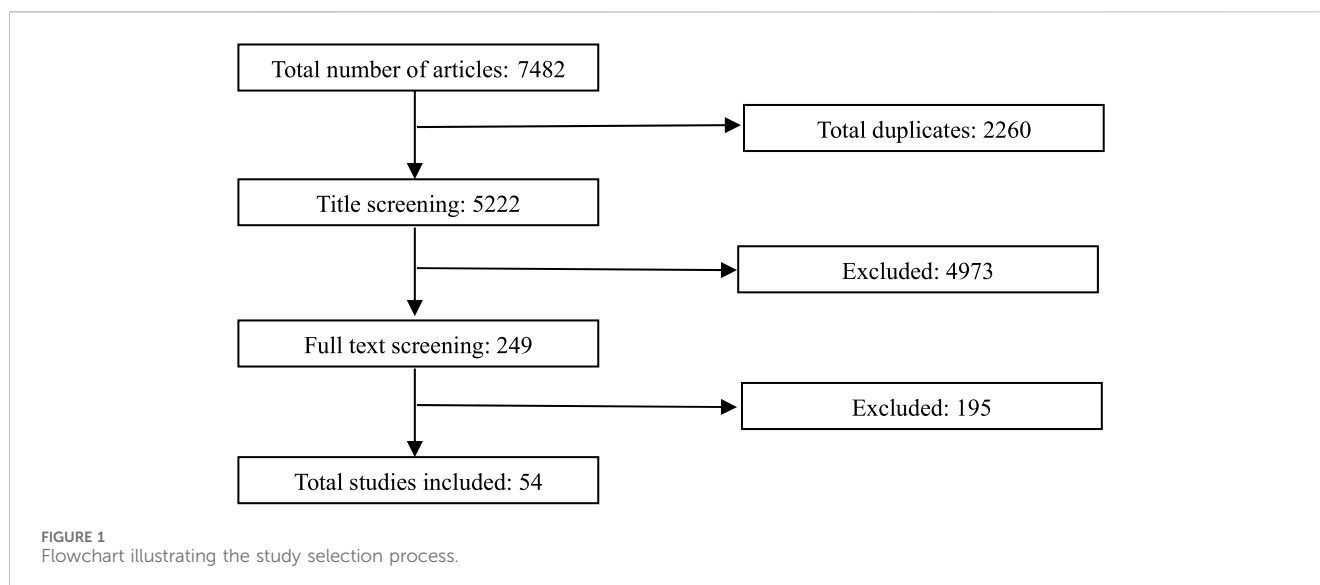
A narrative, structured literature review was conducted between November 2024 and February 2025. The objective was to identify all relevant evidence on clinical uncertainties in ATMPs, with a particular focus on the use and assessment of indirect treatment comparisons (ITCs) and surrogate endpoints within HTA processes.

2.2 Search strategy and information sources

A predefined search strategy was implemented in PubMed and Embase, combining keywords and controlled vocabulary terms related to: Health Technology Assessment, Advanced Therapy Medicinal Products, Indirect Treatment Comparison, and Surrogate Endpoints. The search strategy was developed with the support of a medical librarian to ensure comprehensive coverage of terminology and databases.

The initial search strategy was intentionally broad to capture all relevant methodological discussions in HTA. During title and abstract screening, publications were retained if they explicitly referred to ATMPs or addressed HTA methodological issues relevant to ATMP assessment.

The search was supplemented with grey literature retrieval, including HTA bodies' guidelines, reports from international organizations, and documents issued by relevant professional bodies.



To enhance completeness, a snowballing approach was applied by screening the reference lists of all included publications and relevant reviews.

2.3 Eligibility criteria

Eligible publications were peer-reviewed articles written in English and reporting findings from literature reviews or document analyses. Authoritative grey literature from recognized institutional sources was also included. Clinical studies and non-peer-reviewed documents outside the defined grey literature scope were excluded.

2.4 Study selection

One reviewer screened titles and abstracts followed by full-text assessment. The screening and selection process is summarized in the flowchart (Figure 1). Any methodological approaches outside the primary focus were categorized under innovative trial design if identified during the search.

During title and abstract screening, publications were retained if they referred to ATMPs or to HTA methodological challenges plausibly applicable to ATMP assessments. During full-text screening, inclusion was restricted to articles explicitly addressing how HTA bodies cope with clinical uncertainties related to surrogate endpoints and indirect treatment comparisons in the evaluation of ATMPs.

2.5 Comparative analysis of pharmacoeconomic guidelines

For the comparative assessment of pharmacoeconomic recommendations, we focused on European countries. This geographical scope aligned with the evaluation of clinical trial

evidence available at the time of EMA marketing authorisation, ensuring a consistent regulatory context. Relevant documents were identified using the ISPOR database on pharmacoeconomic guidelines, supplemented by national HTA guidance from selected European countries (ISPOR, 2025).

2.6 Data extraction and synthesis

Data from the included publications were charted according to relevance to clinical uncertainties in ATMPs, with particular attention to indirect treatment comparisons, surrogate endpoints, and innovative trial designs where applicable. Results were synthesized narratively due to heterogeneity in study types, evidence sources, and reporting formats.

3 Results

3.1 Indirect treatment comparison

This section discusses and explores approaches to address clinical uncertainty surrounding the relative effectiveness measurement of an ATMP versus a comparator in a reimbursement context.

3.1.1 RCT as the gold standard

RCTs are considered as the gold standard to measure treatment outcomes. Randomization secures comparability of participants characteristics in both treatment and control or comparator group. This reduces bias and allows to examine the cause-effect relationship between an intervention and an outcome. This principle applies not only to known or observed patient characteristics but also to unknown characteristics, where achieving or even assessing a balanced distribution across groups is not feasible using alternative methods (Collins et al., 2020).

For some ATMPs evidence from RCTs is available (e.g., Luxturna®) (Russell et al., 2017). However RCTs are not always feasible, due to practical or ethical constraints. From an ethical perspective, ATMPs are often developed for life-threatening diseases with limited or no existing treatment options. Requiring patients to participate in RCTs, where some may receive suboptimal standard of care instead of an effective treatment option, raises ethical concerns (Frangoul et al., 2024). For some therapies, no commonly accepted standard of care exists, complicating the choice of comparator (Talib and Boon, 2020). Small patient populations and the complex manufacturing of these therapies further limit trial feasibility (Hanna et al., 2016). In therapies like CAR-T cell treatment, multiple relevant comparators make comprehensive RCTs impractical and resource-intensive.

However, there are currently no clear guidelines for determining the feasibility of conducting an RCT for ATMPs. There is a need for clear guidance to ensure that feasibility assessments are carried out collaboratively with stakeholders at an early stage. If an RCT is deemed not feasible, alternative methods for generating comparative evidence with single-arm trials, such as indirect treatment comparisons, should be considered (Julian et al., 2022).

3.1.2 Characteristics of indirect treatment comparisons

The term indirect treatment comparison (ITC) is mostly used to describe the comparison of the effectiveness of two or more treatments that have not been directly studied in head-to-head trials and rely on a common comparator. However the term can also be used to refer to any study type in which treatment groups from different studies are compared. When RCTs are unavailable and indirect treatment comparisons are used, it is important to recognize that this approach inherently increases uncertainty about the treatment effect, as it cannot ensure the same level of balance for both known and unknown effect modifiers. Therefore, direct comparisons in RCTs with a low risk of bias should be used whenever possible (European Commission, 2024a). Numerous guidelines on ITCs have been issued by regulatory and other authorities such as the Canadian Agency for Drugs and Technologies in Health (Canadian Agency for Drugs and Technologies in Health, 2009), the Food and Drug Administration (U.S. Food and Drug Administration, Center for Drug Evaluation and Research, 2023) and the EU Member State Coordination Group on HTA (European Commission, 2024a), but the main methods for performing an ITC are summarized below including their features, assumptions, (dis)advantages, and the specific circumstances in which they are most appropriate.

All ITC methods rely on the assumption of exchangeability, meaning that substituting patients from one trial with those from another would not change the expected treatment effect. This is assessed through homogeneity, similarity, and consistency (Macabeo et al., 2024a).

- Homogeneity: No variation in relative effectiveness between treatment pairs, ensured by comparable trial designs, patient populations, interventions, and outcomes.
- Similarity: Trials are sufficiently alike regarding effect modifiers, so differences in known or unknown factors that influence outcomes are minimal.

- Consistency: No meaningful differences exist between direct and indirect evidence; significant discrepancies or reversed treatment effects raise concerns.

A number of methods can be used to carry out an ITC (U.S. Food and Drug Administration, Center for Drug Evaluation and Research, 2023; Macabeo et al., 2024a; Tanaka et al., 2024; European Commission, 2024b; Ades et al., 2024).

3.1.2.1 Standard methods

Standard ITC methods, including the Bucher method and network meta-analysis (NMA), enable the estimation of relative treatment effects in the absence of direct head-to-head trials. The Bucher method facilitates indirect comparisons within a simple three-treatment network linked by a common comparator, whereas NMA integrates both direct and indirect evidence across more complex networks, potentially including multiple interventions. Both approaches require the assumptions of homogeneity, similarity, and, in the case of NMA, consistency. These methods can provide robust comparative effectiveness evidence when multiple comparators exist, as exemplified by CAR-T therapies, but they cannot overcome the inherent limitations of single-arm trials.

3.1.2.2 Population-adjusted methods

Population-adjusted ITC (PAICs) methods are used to adjust for imbalances in treatment effect modifiers between studies (Phillippo et al., 2018). Anchored methods rely on a common comparator connecting two or more trials, allowing relative treatment effects to be estimated while preserving the randomised design of the original studies. In contrast, unanchored methods are applied when no common comparator exists, such as in single-arm trials, and estimate treatment effects without any shared reference. Although unanchored ITCs can provide comparative effectiveness evidence in settings where only single-arm data are available, they remain susceptible to bias from unknown or unmeasured effect modifiers, which cannot be accounted for (Hatswell and Sullivan, 2020; Thom et al., 2022).

Propensity score methods use individual patient data (IPD) from both studies to balance measured covariates across treatment groups, thereby reducing confounding. When IPD are available for only one treatment and only aggregate data (AGD) exist for the comparator, Matching-Adjusted Indirect Comparison (MAIC) and Simulated Treatment Comparison (STC) can be applied. MAIC weights patients in the IPD trial to match the baseline characteristics of the aggregate population, while STC uses regression models on the IPD to adjust outcomes.

Consequently, transparent reporting of the rationale for using a PAIC, the selection and clinical justification of treatment-effect modifiers, data sources, balance diagnostics, and sensitivity analyses exploring residual confounding is essential for the interpretation of these analyses in HTA (Phillippo et al., 2018; Signorovitch et al., 2012).

3.1.3 The role of indirect treatment comparison for ATMPs

As of April 2025, 27 ATMPs have obtained market authorisation from the EMA (European Medicines Agency,

TABLE 1 Key characteristics of ATMP studies.

ATMP	Type of ATMP	Non-randomised	Historical control	Surrogate endpoint
Chondroelect	TEP			X
Glybera	GTMP	X		X
MACI	TEP			X
Provenge	CTMP			
Holoclax	TEP	X		X
Imlygic	GTMP			X
Strimvelis	GTMP	X	X	
Zalmoxis	CTMP	X	X	X
Spherox	TEP			X
Alofisel	CTMP			
Yescarta	GTMP	X	X	X
Kymriah	GTMP	X	X	X
Luxturna	GTMP			X
Zynteglo	GTMP	X		X
Zolgensma	GTMP	X	X	
Libmeldy	GTMP	X	X	X
Tecartus	GTMP	X		X
Skysona	GTMP	X	X	
Abecma	GTMP	X	X	X
Breyanzi	GTMP	X		X
Carvykti	GTMP	X		X
Upstaza	GTMP	X	X	
Roctavian	GTMP	X		X
Ebvallo	CTMP	X		X
Hemgenix	GTMP	X		
Casgevy	GTMP	X		X
Durveqtix	GTMP	X		

2025). Of these, only seven were supported by RCTs, whereas 20 relied on single-arm evidence. This evidentiary landscape highlights the increasing reliance on ITCs to inform comparative effectiveness assessments in ATMP evaluations (Table 1).

HTA agencies continue to express caution regarding ITCs, particularly with respect to population-adjusted approaches and most prominently unanchored methods, as these analyses are not randomised and therefore depend on strong, often unverifiable assumptions that can introduce bias. Despite this hesitancy, unanchored population-adjusted ITCs may represent the only viable option when comparative effectiveness from a single-arm trial with historical or external comparators are required. However, their adoption has gained traction because naive comparisons constitute the sole alternative in such contexts (Macabeo et al., 2024b; Katsoulis et al., 2024).

Unanchored population-adjusted ITCs require robust adjustment for all imbalanced prognostic factors and treatment-effect modifiers. In practice, satisfying the assumption of no unobserved confounding is widely considered challenging or even infeasible. The difficulty is amplified by the fact that reporting practices remain highly heterogeneous and frequently incomplete (Truong et al., 2023; Serret-Larmande et al., 2023). Accurate selection and transparent reporting of relevant confounders are crucial. Without these, adjusted estimates may not demonstrate improved validity or precision compared with unadjusted analyses. This situation illustrates the need for strengthened methodological and reporting standards, and recent initiatives have begun to formulate recommendations to address these shortcomings (Freitag et al., 2023).

In addition, population-adjusted ITCs in the ATMP field are conducted almost exclusively in collaboration with pharmaceutical

TABLE 2 HTA guidance on the use and preference for ITC methods.

Country	Standard methods	Population-adjusted methods
The Netherlands	NMA or MTC preferred method of ITC	A single-arm study alone cannot determine relative effectiveness. Therefore, it must be combined with an external control arm for indirect comparison, which introduces significant uncertainty and is never the preferred approach
Denmark	If direct comparison data is unavailable, the company should conduct indirect comparisons using pairwise-adjusted methods, NMA, or other validated approaches	The company should use statistical methods to enhance comparability, adjusting relevant parameters. MAIC or STC can be applied
Belgium	If no direct comparisons are available, indirect comparisons from RCTs can be performed	/
France	The chosen evaluation method should be justified and clearly detailed, an NMA is recommended	A population adjusted indirect comparison method may be used if justified
Germany	Only adjusted indirect comparisons using appropriate common comparators are accepted, including the Bucher method and NMA	MAIC and STC are disapproved due to strong assumptions about unknown effects and reliance on non-testable causal models without a common comparator
Ireland	Adjusted indirect comparison is suitable for two technologies with a common comparator	Single-arm study data should be included only with an anchored comparison
Norway	In the absence of direct efficacy data between the intervention and comparators, indirect comparisons can be performed, including pairwise adjusted indirect comparisons, NMA, or other validated methods	If individual patient data is available for at least one study, methods such as MAIC or STC can be used, provided the relevant conditions for these methods are met
Switzerland	If no direct head-to-head studies are available for two comparators, indirect comparisons can be made using individual studies. The limitations of indirect comparisons due to study differences must be thoroughly discussed	/
Finland	If health effects are assessed through indirect comparisons, proper scientific practices must be followed in both their preparation and reporting	/
Scotland	Elaborate guidelines with a checklist of information for indirect evidence	/
England	When technologies are being compared that have not been evaluated within a single RCT, data from a series of pairwise head-to-head RCTs should be presented together with an NMA if appropriate	Population adjustment methods in connected networks can be used when effect modifiers are imbalanced between trials. The limitations of these methods should be acknowledged, along with an estimate of any potential systematic bias
Portugal	If the evidence base consists only of randomised studies but has disconnected elements, alternative methods can link them. Preferably, this includes expanding the network to incorporate 2nd and 3rd order indirectness	The least preferred approach is the use of unanchored adjustment methods (MAIC, STC) or observational data, as these analyses require careful interpretation

manufacturers. In such analyses, the IPD arm typically represents the sponsor's own therapy, while the aggregated data correspond to a competitor's treatment. Ideally, appropriate adjustment would require IPD for both interventions (Phillippo et al., 2018; Bachy, 2024). However, such data are rarely available. Historically, comparator evidence has mainly consisted of clinical trial control groups. Real-world external controls are increasingly considered preferable, as they enable the construction of contemporary standard-of-care cohorts that better reflect current clinical practice.

Several PAICs have been conducted in Spinal Muscular Atrophy to compare ATMPs against therapies such as nusinersen and risdiplam (Bischof et al., 2021; Ribero et al., 2022). However, differences in trial populations, baseline characteristics, and outcome definitions, along with limited covariate adjustment, leave substantial residual confounding and reduce the validity of these indirect comparisons (Jiang et al., 2023). Similarly, CAR-T therapies are often used in late-line relapsed or refractory settings, and the presence of multiple approved CAR-T products for the same indications further complicates the feasibility of RCTs. Network meta-analyses and PAICs have been performed to compare outcomes across treatments in B-cell non-Hodgkin lymphoma and multiple myeloma (Looka et al., 2025; Liao et al., 2024; Shah

et al., 2022; Jagannath et al., 2021; Alsdorf et al., 2024), but prior studies frequently failed to account for key clinical factors, such as prior treatment history and treatment duration, leading to potentially biased or clinically limited estimates (Cope et al., 2020).

3.1.4 A comparative analysis of pharmacoeconomic guidelines in Europe regarding indirect treatment comparisons

Hence, for this study, we reviewed pharmacoeconomic guidelines from various European countries that are available in English (see Table 2). Nearly all guidelines strongly emphasize a preference for RCTs while acknowledging the use of indirect treatment comparisons. The only justification given for ITC is the lack of direct comparisons, without specific guidelines on when ITC should be considered appropriate. Nearly all guidelines agree that unadjusted indirect comparisons are inappropriate. Within adjusted ITCs, network meta-analysis is most often mentioned for comparing placebo-controlled trials. However, ATMP evaluations often lack placebo-controlled trials and instead rely on single-arm trials.

The German guidelines are particularly stringent, explicitly disapproving of MAIC and STC due to the strong assumptions

required about unknown effects (Institute for Quality and Efficiency in Health Care (IQWiG), 2022). Similarly, though less stringent, the Dutch guidelines state that such methods imply significant uncertainty and are therefore never the preferred approach (Zorginstituut Nederland, 2024). The Portuguese guidelines rank ITC methods hierarchically, placing MAIC and STC at the lowest tier and advising that results from these analyses should be interpreted with caution (INFARMED - National Authority of Medicines and Health Product, 2019).

Other guidelines, such as those from France, Norway, Denmark and Scotland, provide more permissive perspectives. They indicate that population adjusted indirect treatment methods can be considered provided that robust statistical methodologies are employed. Moreover, these guidelines emphasize the need to justify the chosen method, and biases must be discussed and explained.

Overall, there is a pressing need for more comprehensive guidelines addressing single-arm trials, particularly when systematic searches reveal an absence of suitable comparators. The Scottish guidelines stand out by offering clear guidance on the use of indirect treatment comparisons, setting a potential model for addressing these challenges in ATMP reimbursement evaluations.

3.2 Surrogate endpoints

This section discusses and explores approaches to address the use of surrogate endpoints in the assessment of ATMPs for reimbursement purposes.

A clinical endpoint, also referred to as a patient-relevant outcome, is a direct measure of how a patient feels, functions or survives (Biomarkers, Definitions Working Group, 2001). Surrogate endpoints are used to replace and predict a clinical endpoint that cannot be observed in the trial. They can be a biomarker or an intermediate clinical endpoint (Ciani et al., 2017). Whilst there is a preference for clinical endpoints, the need to use surrogate endpoints is recognised (Ciani et al., 2023). Surrogate endpoints are increasingly used in clinical research because they allow earlier assessment of treatment efficacy when long-term outcomes require extended follow-up (Ciani et al., 2021). They are particularly valuable for diseases with small patient populations, as surrogate endpoints can be measured more frequently, allowing trials to demonstrate treatment effects with fewer participants. In severe or life-threatening conditions, their use can also be ethically justified, since waiting for long-term clinical outcomes may place patients at unnecessary risk. EUnetHTA has developed guidelines on the use of surrogate endpoints, which have been adopted by international HTA bodies (Grigore et al., 2020). These guidelines state that the acceptability of a surrogate endpoint is based on biological plausibility and empirical evidence (EUnetHTA, 2015). There are three levels of evidence taken into account for the endpoint-clinical outcome relationship (Fleming and DeMets, 1996). Level one evidence is established by assessing the correlation coefficient between the treatment effect on the surrogate endpoint and the clinical endpoint across multiple RCTs or through meta-analysis. Level two evidence is established by demonstrating a consistent association between the surrogate endpoint and the clinical

endpoint from epidemiological/observational studies. Level three evidence is based solely on biological plausibility, derived from pathophysiological studies or an understanding of the disease process. Furthermore, the validation of surrogate endpoints appears to be disease specific, population specific and technology specific.

The use of surrogate endpoints introduces uncertainty into the assessment of clinical benefit, which can lead to differing incentives for regulators and payers. Regulators may approve therapies to enable early patient access, while payers face uncertainty about true clinical outcomes and adopt more cautious coverage decisions (Dawoud et al., 2021; Bognar et al., 2017). HTA bodies prefer that not only the primary endpoint, but also all other relevant endpoints are clinical endpoints (Pavlovic et al., 2014). Surrogate endpoints are generally considered only when clinical endpoints are unavailable and the surrogate endpoint is validated (Velasco Garrido and Mangiapane, 2009). However, a recent study found that the use of surrogate endpoints was not associated with an increased likelihood of a negative reimbursement decision in cancer drugs (Pinto et al., 2020). In line with this, an analysis of HTA reports and a choice experiment study revealed that HTA bodies' acceptance of surrogate endpoints does not necessarily translate into a greater likelihood of positive coverage decisions (Ciani et al., 2022). In contrast, cost-effectiveness emerged as a significant determinant in these decisions, consistent with findings from earlier research (Ledford, 2025). However, cost-effectiveness analyses frequently involve extrapolating surrogate endpoints to predict clinical outcomes (Taylor and Elston, 2009). Previous systematic reviews have demonstrated that drugs approved on the basis of surrogate endpoints often fail to show a corresponding benefit on definitive clinical endpoints (Wheaton and Bujkiewicz, 2025; Cooper et al., 2020).

3.2.1 The role of surrogate endpoints for ATMPs

As of April 2025, a total of 27 ATMPs have been approved by the European Medicines Agency (European Medicines Agency, 2025). Of these 19 relied on the use of surrogate endpoints (see Table 1).

Due to their curative potential and intended long-term clinical benefits, ATMPs significantly rely on surrogate endpoints to demonstrate efficacy for reimbursement when definitive outcomes, like overall survival, have not yet matured (Ciani et al., 2023). However in the context of ATMPs, it is challenging to meet the prerequisites for the validation of surrogate endpoints as described above (Gladwell et al., 2024). Level one evidence requires multiple RCTs that demonstrate a robust trial-level association between treatment-induced changes in the surrogate endpoint and corresponding changes in the ultimate outcome of interest. Moreover, such trials must be conducted in the same target population, for the same disease, and within the same therapeutic class as the novel intervention. As noted above, ATMPs are a novel technology often developed for diseases with no existing treatment options. Consequently, RCT data are frequently unavailable, and surrogate endpoints validated in other contexts cannot be directly transferred, as their applicability is technology-, disease-, and population-specific. To help address these challenges, new methods are developed such as the causal inference approach with mixed models for longitudinal outcomes, to establish robust validation using limited single-trial data (Roberts et al., 2023).

Another example of such a novel methodology is the use of Bayesian Evidence Synthesis methods to formally integrate real-world evidence with trial data for a more robust surrogate endpoint validation (Wheaton et al., 2023). For oncology, there is a substantial literature on the relationship between progression-free survival and overall survival, but this discussion is beyond the scope of the current literature review. A number of factors affect this relationship and should consider both tumour-type and other relevant clinical and patient factors (Assouline et al., 2022; Sola-Morales et al., 2019; Lux et al., 2021; Campbell et al., 2019; Holstein et al., 2019).

In addition to the reliance on surrogate endpoints, HTA bodies are also required to extrapolate clinical data to estimate long-term outcomes such as overall survival and sustained clinical benefit, which are rarely fully observed within the limited duration of clinical trials (Hardy and Hughes, 2022). This extrapolation is particularly important for cost-effectiveness analyses, as projected lifetime health gains strongly influence value assessments and reimbursement decisions. Multiple modelling techniques exist to extrapolate clinical data, with techniques having different characteristics and making different assumptions (Latimer, 2013). Nevertheless, the assumptions underpinning extrapolation models introduce significant uncertainty, and even small variations in survival curve modelling can lead to markedly different incremental cost-effectiveness ratios (Tai et al., 2021). One proposed approach to managing this persistent uncertainty is to embed reassessment into the HTA and reimbursement process. Reassessing a technology at predefined intervals aligns well with the long-term evidence generation pathways of ATMPs, enabling decisions to be updated as more mature data become available. Such a life-cycle approach would ensure that products with uncertain durability of benefit are not prematurely locked into reimbursement conditions (Versteeg et al., 2024). To mitigate these challenges, it would be helpful if reimbursement agencies clarify which factors they take into account when judging the maturity of clinical data and to offer guidance on validated modelling practices. Extensive sensitivity analyses, and the integration of real-world evidence are increasingly emphasized in HTA evaluations of ATMPs.

3.2.2 A comparative analysis of pharmacoeconomic guidelines in Europe regarding surrogate endpoints

In this study, we reviewed pharmacoeconomic guidelines from various European countries, limited to documents available in English. Among the countries analyzed, only three provided explicit definitions of surrogate endpoints. Portugal (Vinhas et al., 2021) defines surrogate endpoints exclusively as biomarkers. In contrast, France (Haute Autorité de Santé, 2020) and Ireland (Health Information and Quality Authority, 2018) classify them as intermediate endpoints. Ireland requires these endpoints to be objectively measurable, meaning that the outcomes must be quantified directly through standardized instruments or tests and should produce values that are independent of assessor interpretation. France, on the other hand, also accepts observational measures, meaning that outcomes can be based on systematic clinician or researcher observation of patient signs,

symptoms, or behaviors, where structured assessments are applied but some degree of subjective judgment remains unavoidable.

As previously noted, the validation and transferability of surrogate endpoints are particularly important in the context of ATMPs. Portugal and Germany (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 2023) are the only countries that explicitly state that transferability should be assessed based on the specific disease, population, and technology. However, the Portuguese guidelines make an exception for novel health technologies: in such cases, surrogate endpoints may be based on evidence from drugs with different mechanisms of action, and are considered acceptable only when no therapeutic alternatives exist or when treating serious conditions. England and Wales (National Institute for Health and Care Excellence, 2022), indicate that transferability should be assessed with respect to both the technology and the disease context, while France limits this consideration to disease specificity alone. Other countries do not explicitly address the transferability of surrogate endpoints.

Most countries (see Table 3) emphasize the need for a demonstrated causal relationship between the surrogate and clinical endpoint. Three countries refer to a staged validation process. Germany, for instance, only accepts validation based on meta-analyses, whereas alternative methods may be considered by other countries but require strong justification. Portugal is the only country to specify quantitative thresholds, requiring a correlation coefficient of 0.85 or 0.955 between the surrogate and clinical endpoint. Lastly, Scotland (Healthcare Improvement Scotland, 2022) and England and Wales note that validation requirements may vary in the context of orphan or rare diseases.

3.3 Innovative trial design

Innovative clinical trial designs are increasingly recognized as essential for addressing key challenges in ATMP evidence generation, particularly the frequent reliance on single-arm studies and surrogate endpoints. These designs can maximize limited patient populations, integrate real-world data, and support earlier assessment of clinical outcomes. Below, we outline several innovative trial types and illustrate their potential contributions to ATMP development.

An adaptive trial allows pre-planned modifications based on interim data, such as adjusting sample size, modifying randomization ratios, adding or dropping treatment arms, or stopping early for success or futility (Bhatt and Mehta, 2016). This methodology is used in different innovative trial designs mentioned below.

A master protocol tests multiple treatments and/or subpopulations within a single overarching protocol, eliminating the need for separate studies. There are three main types: umbrella trials, which test multiple treatments for a single disease; basket trials, which test a single treatment across multiple diseases or subtypes; and platform trials, which allow multiple treatments to be evaluated adaptively over time (Woodcock and LaVange, 2017). This design is particularly useful for ATMPs because it enables efficient evaluation in small and heterogeneous populations, allows simultaneous testing of multiple strategies, and adapts to emerging

TABLE 3 HTA guidance on the use and preference for surrogate endpoints.

Country	Definition	Validity	Transferability
Germany	/	Validation requires staged validation, preferably meta-analysis, alternatives are only considered in justified exceptions. If a surrogate endpoint cannot be validated conclusively, the surrogate threshold effect concept can be used	Transferability is population-specific, technology-specific and disease specific
Norway	/	Validation requires a causal correlation	/
France	A biomarker or intermediate outcome that can be considered a relevant substitute for a clinical endpoint	Validation requires a causal correlation	/
Denmark	/	Validation requires a causal correlation	/
Ireland	An objectively measured endpoint expected to predict clinical benefit or harm, based on epidemiologic, pathophysiologic, therapeutic, or other scientific evidence	Validation requires a causal correlation and a comparable effect size	/
Austria	/	Validation requires a causal correlation	/
Portugal	A biomarker intended to replace a clinical outcome measure	Validation requires staged validation. Correlation coefficient values between 0.85 and 0.955 are given. If there is not a high correlation, the surrogate threshold effect can still be used	Transferability is population-specific, technology-specific and disease specific. To validate a surrogate endpoint for a new technology, use evidence from other drugs for the same indication only if no alternatives exist or added benefit is expected
England and Wales	/	Validation requires staged validation. For some rare diseases evidence levels may be lower	Transferability is population-specific and technology-specific
Scotland	/	Validation requires a causal correlation	/

data. For example, the GLO-BNHL study is a global, adaptive platform trial evaluating several novel agents, including CAR-T therapies, for pediatric and young adult patients with relapsed or refractory B-cell Non-Hodgkin lymphoma, allowing ineffective treatments to be dropped while promising ones continue.

Seamless trials combine multiple phases within a single protocol, often reusing patients and infrastructure. Inferentially seamless designs incorporate early-phase data into later analyses, increasing efficiency but requiring complex statistical planning, whereas operationally seamless designs maintain continuity across phases but rely only on confirmatory-stage data for final evaluation (Cuffe et al., 2014). The UX701 trial for Wilson disease, an operationally seamless Phase I/II/III study, exemplifies how this approach can accelerate development in rare diseases with recruitment challenges.

The N-of-1 trial, also known as a single-patient trial or personalized trial, is a research design where a single individual is the only participant and serves as their own control. In this design a patient receives an intervention alternating with a placebo or a different intervention, over a series of pre-defined periods (Samuel et al., 2023). The order of interventions is typically randomised and blinded to both the patient and the clinician to minimize bias. These trials are ideal for chronic stable health conditions. For a single-administration therapy like ATMPs, the traditional alternating N-of-1 design can be used. Since a classic crossover isn't feasible, the trial establishes a patient's baseline before administering the therapy, followed by long-term monitoring of their efficacy and safety (Kim-McManus et al., 2024). This innovative design has been used with Milasen[®], a patient-customized oligonucleotide therapy for Batten's disease which demonstrates similarities to ATMPs despite not being strictly classified as one (Kim et al., 2019).

Pragmatic clinical trials are designed to evaluate the effectiveness of interventions under real-world conditions. These trials often rely on a broad range of outcomes derived from real-world data (RWD) (Ford and Norrie, 2016). For ATMPs, which frequently enter the market with limited clinical trial data, post-marketing RWE is particularly valuable in addressing evidence gaps and complementing initial trial findings by providing crucial long-term safety and effectiveness data from diverse patient groups in real-world settings. In the case of Strimvelis[®], an analysis of treated patients, including those in the post-marketing phase, confirmed durable efficacy and a favorable risk-benefit profile (Migliavacca et al., 2024).

Innovative trial designs provide practical solutions to the limitations of single-arm studies and surrogate endpoints in ATMP development. Master protocols can integrate multiple sub-studies and shared control data, reducing reliance on independent comparators. Seamless and adaptive trials allow iterative modification of endpoints and sample sizes, supporting early evaluation and refinement of surrogate endpoints. Pragmatic trials leverage real-world data to provide context for comparative effectiveness when conventional randomization is not feasible. By combining these approaches, developers can generate more robust, interpretable evidence, contextualize surrogate endpoints, and strengthen the overall reliability of clinical and economic assessments for ATMPs.

4 Discussion

ATMPs are transforming the landscape of medical treatment by offering curative potential for previously intractable diseases. In this article, we reviewed the literature and analyzed European

pharmacoeconomic guidelines to examine how key methodological clinical challenges for ATMPs such as the absence of appropriate comparators and the uncertainty of long-term clinical outcomes can be addressed in the context of reimbursement evaluation.

Most ATMPs gain market access based on single-arm studies (20/27), which frequently rely on surrogate endpoints (19/27). There is a clear need for stronger, more explicit guidance on assessing the feasibility of conducting an RCT, particularly when robust comparative effectiveness evidence is required. Currently, no standardized criteria exist, and feasibility evaluations are not consistently performed in collaboration with stakeholders at an early stage. When an RCT is deemed infeasible, there should be a well-supported rationale explaining why, including why innovative trial designs cannot address the identified challenges. In the absence of randomization, comparative evidence often relies on indirect ITCs, typically unanchored PAICs due to single-arm trials. However, without a randomised comparator, confounding factors may remain unevenly distributed, introducing bias. Recent research has shown considerable inconsistency in reporting the rationale for selecting effect modifiers and how balance between treatment populations was achieved. Ensuring methodological rigor and transparency in these steps is crucial to provide a reliable picture of comparative effectiveness. Pharmacoeconomic guidelines generally prioritize RCT evidence and accept ITCs only when direct comparisons are not feasible, yet they offer little clarity on when ITCs are appropriate. Perceptions of population-adjusted methods such as MAIC and STC also vary widely: many guidelines remain cautious due to the significant uncertainty these methods entail, while others allow greater flexibility if methodological choices are justified and potential biases transparently addressed.

The potentially curative nature of ATMPs poses significant challenges to traditional HTA frameworks, often necessitating reliance on surrogate endpoints in the absence of long-term clinical data. The validation of such endpoints typically depends on evidence from RCTs, and their applicability is highly context-specific. While rigorous validation remains essential, current methodological frameworks are frequently insufficient to address the unique requirements of ATMPs. Moreover, most pharmacoeconomic guidelines do not provide explicit criteria for determining acceptable levels of surrogate endpoint validation or for managing situations in which RCT-based validation is infeasible and transferability conditions are not met. This uncertainty is further compounded by the difficulty of extrapolating short-term surrogate endpoints, outcomes should be left out to long-term clinical and economic impacts.

Recent regulatory developments illustrate alternative approaches to evidentiary uncertainty. The FDA's *Plausible Mechanism Pathway* allows early clinical evidence to be considered when supported by a well-substantiated biological rationale, even when conventional surrogate endpoint validation through randomised trials is not feasible (Prasad and Makary, 2025). While developed in a regulatory context, this approach highlights considerations that are also relevant for HTA of ATMPs, particularly in settings characterized by small populations, single-arm studies, and immature outcome data, and underscores the importance of transparency and post-marketing evidence generation. To mitigate these challenges, innovative trial designs offer a structured approach to improving evidence generation. Master protocols, seamless trials, and pragmatic designs can maximize the information obtained from limited patient populations, incorporate real-world evidence, and support iterative evaluation of surrogate endpoints. While they cannot replace

randomised comparators, these approaches can enhance the interpretability and contextual relevance of single-arm studies. Formally incorporating such designs into HTA frameworks and outcomes-based MEAs could help reduce uncertainty in reimbursement decisions, even in the absence of conventional RCTs.

This lack of standardization risks delaying patient access to potentially transformative treatments. To mitigate this, early and structured dialogue between developers and payers is essential to reduce evidentiary uncertainty and streamline the reimbursement process. An example of such collaborative efforts is the newly established Joint Clinical Assessment (JCA) framework under the EU HTA Regulation (Regulation EU, 2021). Building on the groundwork laid by EUnetHTA, the JCA provides a formal mechanism for cross-country cooperation through joint scientific consultations, aiming to ensure more consistent, transparent, and efficient HTAs across Europe. Possibly supported by bodies such as the JCA, it will become increasingly important for HTA agencies to establish clear guidance on the use of PAICs and surrogate endpoints. New ATMPs will continue to face challenges related to the lack of suitable comparators and the absence of long-term clinical data. It is therefore crucial that HTA agencies acknowledge these limitations and provide explicit criteria for when ITCs are appropriate, as well as for the acceptable use and extrapolation of surrogate endpoints.

Given the frequent immaturity of clinical data at market entry, post-marketing evidence generation will continue to play a pivotal role. Here, RWE becomes indispensable (Dang, 2023). Initiatives like the European Health Data Space have the potential to facilitate robust RWE collection across jurisdictions, provided that challenges related to data interoperability, quality, and governance are adequately addressed (Regulation EU, 2025).

Finally, Managed Entry Agreements (MEAs) will likely remain a key tool for managing uncertainty (Vreman et al., 2020). Outcomes-based MEAs, which link reimbursement to real-world effectiveness, are particularly relevant (Greco et al., 2025). However, their feasibility is limited by the same challenges that affect initial evidence generation, including single-arm trials, lack of comparators, and reliance on surrogate endpoints. Without robust comparators and validated endpoints, it is difficult to attribute observed outcomes to the therapy and to predict long-term benefit. Successful implementation also requires strong real-world evidence infrastructure, including standardized data collection, transparent registries, and consensus on meaningful endpoints. Without these elements, operational complexity may outweigh benefits, delaying access and threatening sustainable reimbursement (Dabbous et al., 2020).

It is important to acknowledge several limitations of this study. The literature was not identified through a systematic search and the synthesis was narrative in nature, which may have resulted in the omission of relevant guidelines. In addition, the selection of materials was performed by a single reviewer, potentially introducing selection bias. Furthermore, the inclusion was restricted to guidelines available in English, which may have limited the comprehensiveness and representativeness of the findings.

5 Conclusion

Our study has found that there is a need to further optimize methodologies to handle evidence from single-arm trials and to

project long-term outcomes, both of which are common in ATMP development. Nonetheless, a certain level of clinical uncertainty will always be inherent in the reimbursement evaluation of these innovative therapies. This makes early and constructive dialogue between industry and payers essential. The task for modern healthcare systems is to manage these clinical uncertainties associated with ATMPs in a way that ensures their equitable and timely patient access without creating new barriers to care.

Author contributions

LD: Conceptualization, Writing – original draft, Writing – review and editing, Methodology, Data curation, Formal Analysis. IH: Writing – review and editing. WV: Writing – review and editing. SS: Conceptualization, Methodology, Writing – review and editing, Supervision.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the Research Foundation of Flanders (FWO, Fonds voor Wetenschappelijk Onderzoek), the grant number is G0A8Y24N.

Acknowledgements

We wish to thank Michiel Vanhaeren for providing his insights and feedback during the conduct and finalization of the study.

References

- Ades, A. E., Welton, N. J., Dias, S., Phillippo, D. M., and Caldwell, D. M. (2024). Twenty years of network meta-analysis: continuing controversies and recent developments. *Res. Synth. Methods* 15 (5), 702–727. doi:10.1002/jrsm.1700
- Adorf, W., Diels, J., Ghilotti, F., Mendes, J., Hernando, T., Cost, P., et al. (2024). Efficacy of CARVYKTI in CARTITUDE-4 versus other conventional treatment regimens for lenalidomide-refractory multiple myeloma using inverse probability of treatment weighting. *J. Comp. Eff. Res.* 13 (9), e240080. doi:10.57264/ceer-2024-0080
- Assouline, S., Wiesinger, A., Spooner, C., Jovanovic, J., and Schlueter, M. (2022). Validity of event-free survival as a surrogate endpoint in haematological malignancy: review of the literature and health technology assessments. *Crit. Rev. Oncol. Hematol.* 175, 103711. doi:10.1016/j.critrevonc.2022.103711
- Bachy, E. (2024). Indirect treatment comparisons: how to MAIC it right? *Haematologica* 109 (7), 2032–2034. doi:10.3324/haematol.2023.284534
- Bhatt, D. L., and Mehta, C. (2016). Adaptive designs for clinical trials. *N. Engl. J. Med.* 375 (1), 65–74. doi:10.1056/NEJMra1510061
- Biomarkers, Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69 (3), 89–95. doi:10.1067/mcp.2001.113989
- Bischof, M., Lorenzi, M., Lee, J., Druyts, E., Balijepalli, C., and Dabbous, O. (2021). Matching-adjusted indirect treatment comparison of onasemnogene abeparovex and nusinersen for the treatment of symptomatic patients with spinal muscular atrophy type 1. *Curr. Med. Res. Opin.* 37 (10), 1719–1730. doi:10.1080/03007995.2021.1947216
- Bognar, K., Romley, J. A., Bae, J. P., Murray, J., Chou, J. W., and Lakdawalla, D. N. (2017). The role of imperfect surrogate endpoint information in drug approval and reimbursement decisions. *J. Health Econ.* 51, 1–12. doi:10.1016/j.jhealeco.2016.12.001
- Campbell, L., Panitsas, F., Basu, S., Anyanwu, F., Lee, S., Ferry, B., et al. (2019). Serological normalisation as a surrogate marker for minimal residual disease negativity in multiple myeloma. *Br. J. Haematol.* 185 (4), 775–778. doi:10.1111/bjh.15615
- Canadian Agency for Drugs and Technologies in Health (2009). *Indirect treatment comparison user guide*. Ottawa, Canada.
- Cargill, S. S., Eidsvik, A., and Lamm, M. (2021). Updating the ethical guidance for gene and cell therapy research participation. *Mol. Ther.* 29 (8), 2394–2395. doi:10.1016/j.ymthe.2021.07.009
- Ciani, O., Buyse, M., Drummond, M., Rasi, G., Saad, E. D., and Taylor, R. S. (2017). Time to review the role of surrogate end points in health policy: state of the art and the way forward. *Value Health* 20 (3), 487–495. doi:10.1016/j.jval.2016.10.011
- Ciani, O., Grigore, B., Blommestein, H., de Groot, S., Mollenkamp, M., Rabbe, S., et al. (2021). Validity of surrogate endpoints and their impact on coverage recommendations: a retrospective analysis across international health technology assessment agencies. *Med. Decis. Mak.* 41 (4), 439–452. doi:10.1177/0272989X21994553
- Ciani, O., Grigore, B., and Taylor, R. S. (2022). Development of a framework and decision tool for the evaluation of health technologies based on surrogate endpoint evidence. *Health Econ.* 31 (1), 44–72. doi:10.1002/hec.4524
- Ciani, O., Manyara, A. M., Davies, P., Stewart, D., Weir, C. J., Young, A. E., et al. (2023). A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. *EClinicalMedicine* 65, 102283. doi:10.1016/j.eclim.2023.102283
- Collins, R., Bowman, L., Landray, M., and Peto, R. (2020). The magic of randomization versus the myth of real-world evidence. *N. Engl. J. Med.* 382 (7), 674–678. doi:10.1056/NEJMs1901642
- Cooper, K., Tappenden, P., Cantrell, A., and Ennis, K. (2020). A systematic review of meta-analyses assessing the validity of tumour response endpoints as surrogates for progression-free or overall survival in cancer. *Br. J. Cancer* 123 (11), 1686–1696. doi:10.1038/s41416-020-01050-w
- Cope, S., Toor, K., Popoff, E., Fonseca, R., Landgren, O., Mateos, M. V., et al. (2020). Critical appraisal of published indirect comparisons and network meta-analyses of

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author SS declared that they were an editorial board member of *Frontiers* at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. As non-native English speakers, we used ChatGPT to assist with improving the clarity of our writing, specifically for spelling and grammar. No part of the analysis, results, or conclusions was generated by the tool.

Any alternative text (alt text) provided alongside figures in this article has been generated by *Frontiers* with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.

- competing interventions for multiple myeloma. *Value Health* 23 (4), 441–450. doi:10.1016/j.jval.2019.11.003
- Cuffe, R. L., Lawrence, D., Stone, A., and Vandemeulebroecke, M. (2014). When is a seamless study desirable? Case studies from different pharmaceutical sponsors. *Pharm. Stat.* 13 (4), 229–237. doi:10.1002/pst.1622
- Dabbous, M., Chachoua, L., Caban, A., and Toumi, M. (2020). Managed entry agreements: policy analysis from the European perspective. *Value Health* 23 (4), 425–433. doi:10.1016/j.jval.2019.12.008
- Dang, A. (2023). Real-world evidence: a primer. *Pharm. Med.* 37 (1), 25–36. doi:10.1007/s40290-022-00456-6
- Dawoud, D., Naci, H., Ciani, O., and Bujkiewicz, S. (2021). Raising the bar for using surrogate endpoints in drug regulation and health technology assessment. *BMJ* 374, n2191. doi:10.1136/bmj.n2191
- Day, S., Jonker, A. H., Lau, L. P. L., Hilgers, R. D., Irony, I., Larsson, K., et al. (2018). Recommendations for the design of small population clinical trials. *Orphanet J. Rare Dis.* 13 (1), 195. doi:10.1186/s13023-018-0931-2
- Detela, G., and Lodge, A. (2019). EU regulatory pathways for ATMPs: Standard, accelerated and adaptive pathways to marketing authorisation. *Mol. Ther. Methods Clin. Dev.* 13, 205–232. doi:10.1016/j.omtm.2019.01.010
- EUNETHTA (2015). Guideline on the use of surrogate endpoints in health technology assessment.
- European Commission (2024a). “Methodological guideline for quantitative evidence synthesis: direct and indirect comparisons,” in *CG MSCGoHTA*.
- European Commission (2024b). Methodological guideline for quantitative evidence synthesis: direct and indirect comparisons. Member state coordination group on health technology assessment.
- European Medicines Agency (2025). CAT quarterly highlights and approved ATMPs May 2025. Available online at: https://www.ema.europa.eu/en/documents/committee-report/cat-quarterly-highlights-approved-atmps-may-2025_en.pdf.
- European Parliament and Council of the European Union (2007). Regulation (EC) No 1394/2007 of 13 November 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No 726/2004. *J. European Union L324*, 121–137. Available online at: <https://eur-lex.europa.eu/eli/reg/2007/1394/oj/eng>.
- Fleming, T. R., and DeMets, D. L. (1996). Surrogate end points in clinical trials: are we being misled? *Ann. Intern. Med.* 125 (7), 605–613. doi:10.7326/0003-4819-125-7-199610010-00011
- Fontrier, A. M., Visintin, E., and Kanavos, P. (2022). Similarities and differences in health technology assessment systems and implications for coverage decisions: evidence from 32 countries. *Pharmacoecon Open* 6 (3), 315–328. doi:10.1007/s41669-021-00311-5
- Ford, I., and Norrie, J. (2016). Pragmatic trials. *N. Engl. J. Med.* 375 (5), 454–463. doi:10.1056/NEJMra1510059
- Frangoul, H., Locatelli, F., Sharma, A., Bhatia, M., Mapara, M., Molinari, L., et al. (2024). Exagamglogene autotemcel for severe sickle cell disease. *N. Engl. J. Med.* 390 (18), 1649–1662. doi:10.1056/NEJMoa2309676
- Freitag, A., Gurskyte, L., and Sarri, G. (2023). Increasing transparency in indirect treatment comparisons: is selecting effect modifiers the missing part of the puzzle? A review of methodological approaches and critical considerations. *J. Comp. Eff. Res.* 12 (10), e230046. doi:10.57264/ce-2023-0046
- Friede, T., Posch, M., Zohar, S., Alberti, C., Benda, N., Comets, E., et al. (2018). Recent advances in methodology for clinical trials in small populations: the InSPIRe project. *Orphanet J. Rare Dis.* 13 (1), 186. doi:10.1186/s13023-018-0919-y
- Gladwell, D., Ciani, O., Parnaby, A., and Palmer, S. (2024). Surrogacy and the valuation of ATMPs: taking our place in the evidence generation/assessment continuum. *Pharmacoeconomics* 42 (2), 137–144. doi:10.1007/s40273-023-01334-y
- Gozzo, L., Romano, G. L., Romano, F., Brancati, S., Longo, L., Vitale, D. C., et al. (2021). Health technology assessment of advanced therapy medicinal products: Comparison among 3 European countries. *Front. Pharmacol.* 12, 755052. doi:10.3389/fphar.2021.755052
- Greco, A., Frederix, G. W. J., Hooft, L., and Ten Ham, R. M. T. (2025). A systematic review of challenges and opportunities in the implementation of managed entry agreements for advanced therapy medicinal products. *Clin. Ther.* 47 (2), e16–e26. doi:10.1016/j.clinthera.2024.11.019
- Grigore, B., Ciani, O., Dams, F., Federici, C., de Groot, S., Mollenkamp, M., et al. (2020). Surrogate endpoints in health technology assessment: an international review of methodological guidelines. *Pharmacoeconomics* 38 (10), 1055–1070. doi:10.1007/s40273-020-00935-1
- Hanna, E., Remuzat, C., Auquier, P., and Toumi, M. (2016). Advanced therapy medicinal products: current and future perspectives. *J. Mark. Access Health Policy* 4, 31036. doi:10.3402/jmahp.v4.31036
- Hardy, W. A. S., and Hughes, D. A. (2022). Methods for extrapolating survival analyses for the economic evaluation of advanced therapy medicinal products. *Hum. Gene Ther.* 33 (17–18), 845–856. doi:10.1089/hum.2022.056
- Harricharan, S., Gurjar, K., Nguyen, K., and Forsythe, A. (2021). PCN172 evaluation of real-world evidence (RWE) used within health technology Assessments (HTA) in oncology with pivotal single-arm trials: a comparative study of six HTA agencies. *Value Health* 24, S51. doi:10.1016/j.jval.2021.04.263
- Hatswell, A. J., and Sullivan, W. G. (2020). Creating historical controls using data from a previous line of treatment - two non-standard approaches. *Stat. Methods Med. Res.* 29 (6), 1563–1572. doi:10.1177/0962280219826609
- Haute Autorité de Santé (2020). Principles of medicinal product assessments and appraisal for reimbursement purposes.
- Health Information and Quality Authority (2018). Guidelines for evaluating the clinical effectiveness of health technologies in Ireland.
- Healthcare Improvement Scotland (2022). Guidance to submitting companies for completion of new product assessment form.
- Holstein, S. A., Suman, V. J., and McCarthy, P. L. (2019). Should overall survival remain an endpoint for multiple myeloma trials? *Curr. Hematol. Malig. Rep.* 14 (1), 31–38. doi:10.1007/s11899-019-0495-9
- Iglesias-Lopez, C., Agusti, A., Vallano, A., and Obach, M. (2021a). Methodological characteristics of clinical trials supporting the marketing authorisation of advanced therapies in the European Union. *Front. Pharmacol.* 12, 773712. doi:10.3389/fphar.2021.773712
- Iglesias-Lopez, C., Agusti, A., Vallano, A., and Obach, M. (2021b). Current landscape of clinical development and approval of advanced therapies. *Mol. Ther. Methods Clin. Dev.* 23, 606–618. doi:10.1016/j.omtm.2021.11.003
- INFARMED - National Authority of Medicines and Health Product (2019). Methodological guideline for economic evaluation studies of health technologies.
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2023). *General Methods*.
- Institute for Quality and Efficiency in Health Care (IQWiG) (2022). *General methods*.
- ISPOR (2025). Pharmacoeconomic guidelines around the world. Available online at: <https://web.archive.org/web/20250618211940/https://www.ispor.org/heor-resources/more-heor-resources/pharmacoeconomic-guidelines>.
- Jagannath, S., Lin, Y., Goldschmidt, H., Reece, D., Nooka, A., Senin, A., et al. (2021). KarMMa-RW: comparison of ibrutinib with real-world outcomes in relapsed and refractory multiple myeloma. *Blood Cancer J.* 11 (6), 116. doi:10.1038/s41408-021-00507-2
- Jiang, T., Youn, B., Paradis, A. D., Beckerman, R., Barnieh, L., and Johnson, N. B. (2023). A critical appraisal of matching-adjusted indirect comparisons in spinal muscular atrophy. *Adv. Ther.* 40 (7), 2985–3005. doi:10.1007/s12325-023-02520-2
- Julian, E., Pavlovic, M., Sola-Morales, O., Gianfrate, F., Toumi, M., Bucher, H. C., et al. (2022). Shaping a research agenda to ensure a successful European health technology assessment: insights generated during the inaugural convention of the European access academy. *Health Econ. Rev.* 12 (1), 54. doi:10.1186/s13561-022-00402-x
- Katsoulis, I., Graham, A., Thompson, A., Gharibian, N., Pawar, V., Khurana, V., et al. (2024). Perceptions of indirect treatment comparisons as an evidence base in oncology decision-making: results of an international survey of health technology assessment and payer decision-makers. *J. Comp. Eff. Res.* 13 (11), e240040. doi:10.57264/ce-2024-0040
- Kim, J., Hu, C., Moufawad El Achkar, C., Black, L. E., Douville, J., Larson, A., et al. (2019). Patient-customized oligonucleotide therapy for a rare genetic disease. *N. Engl. J. Med.* 381 (17), 1644–1652. doi:10.1056/NEJMoa1813279
- Kim-McManus, O., Gleeson, J. G., Mignon, L., Smith Fine, A., Yan, W., Nolen, N., et al. (2024). A framework for N-of-1 trials of individualized gene-targeted therapies for genetic diseases. *Nat. Commun.* 15 (1), 9802. doi:10.1038/s41467-024-54077-5
- Latimer, N. R. (2013). Survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med. Decis. Mak.* 33 (6), 743–754. doi:10.1177/0272989X124272398
- Ledford, H. (2025). World’s first personalized CRISPR therapy given to baby with genetic disease. *Nature*. doi:10.1038/d41586-025-01496-z
- Liao, C., Zeng, L., Lu, S., Zheng, S., Guo, B., Ke, Q., et al. (2024). Comparison of the efficacy and safety of Axi-Cel and tisa-Cel based on meta-analysis. *J. Cancer* 15 (17), 5729–5741. doi:10.7150/jca.99427
- Looka, A., Qualls, D. A., Matthews, D., Redd, R. A., Sakellis, C., Duffy, C., et al. (2025). A real-world comparison of commercial-use axicabtagene ciloleucel and lisocabtagene maraleucel in large B-cell lymphoma. *Blood Adv.* 9 (3), 455–462. doi:10.1182/bloodadvances.2024012992
- Lux, M. P., Ciani, O., Dunlop, W. C. N., Ferris, A., and Friedlander, M. (2021). The impasse on overall survival in oncology reimbursement decision-making: how can we resolve this? *Cancer Manag. Res.* 13, 8457–8471. doi:10.2147/CMAR.S328058
- Macabeo, B., Quenechdu, A., Aballea, S., Francois, C., Boyer, L., and Laramée, P. (2024a). Methods for indirect treatment comparison: results from a systematic literature review. *J. Mark. Access Health Policy* 12 (2), 58–80. doi:10.3390/jmahp12020006
- Macabeo, B., Rotrou, T., Millier, A., François, C., and Laramée, P. (2024b). The acceptance of indirect treatment comparison methods in oncology by health technology assessment agencies in England, France, Germany, Italy, and Spain. *Pharmacoecon Open* 8 (1), 5–18. doi:10.1007/s41669-023-00455-6
- Migliavacca, M., Barzaghi, F., Fossati, C., Rancoita, P. M. V., Gabaldo, M., Dionisio, F., et al. (2024). Long-term and real-world safety and efficacy of retroviral gene therapy for adenosine deaminase deficiency. *Nat. Med.* 30 (2), 488–497. doi:10.1038/s41591-023-02789-4

- National Institute for Health and Care Excellence (2022). NICE health technology evaluations: the manual.
- Patel, D., Grimson, F., Mihaylova, E., Wagner, P., Warren, J., van Engen, A., et al. (2021). Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health* 24 (8), 1118–1125. doi:10.1016/j.jval.2021.01.015
- Pavlovic, M., Teljeur, C., Wieseler, B., Klemp, M., Cleemput, I., and Neyt, M. (2014). Endpoints for relative effectiveness assessment (REA) of pharmaceuticals. *Int. J. Technol. Assess. Health Care* 30 (5), 508–513. doi:10.1017/S0266462314000592
- Phillippo, D. M., Ades, A. E., Dias, S., Palmer, S., Abrams, K. R., and Welton, N. J. (2018). Methods for population-adjusted indirect comparisons in health technology appraisal. *Med. Decis. Mak.* 38 (2), 200–211. doi:10.1177/0272989X17725740
- Pinho-Gomes, A. C., and Cairns, J. (2022). Evaluation of advanced therapy medicinal products by the national institute for health and care excellence (NICE): an updated review. *Pharmacoecon Open* 6 (2), 147–167. doi:10.1007/s41669-021-00295-2
- Pinto, A., Naci, H., Neer, E., and Mossialos, E. (2020). Association between the use of surrogate measures in pivotal trials and health technology assessment decisions: a retrospective analysis of NICE and CADTH reviews of cancer drugs. *Value Health* 23 (3), 319–327. doi:10.1016/j.jval.2019.10.010
- Prasad, V., and Makary, M. A. (2025). FDA's new plausible mechanism pathway. *N. Engl. J. Med.* 393 (23), 2365–2367. doi:10.1056/NEJMs2512695
- Qiu, T., Pochopien, M., Liang, S., Saal, G., Paterak, E., Janik, J., et al. (2022). Gene therapy evidence generation and economic analysis: pragmatic considerations to facilitate fit-for-purpose health technology assessment. *Front. Public Health* 10, 773629. doi:10.3389/fpubh.2022.773629
- Regulation (EU) (2021). 2021/2282 of the European parliament and of the council of 15 December 2021 on health technology assessment and amending directive 2011/24/EU.
- Regulation (EU) (2025). 2025/327 of the European parliament and of the council of 11 February 2025 on the European health data space and amending directive 2011/24/EU and regulation (EU) 2024/2847.
- Ribero, V. A., Daigl, M., Marti, Y., Gorni, K., Evans, R., Scott, D. A., et al. (2022). How does risdiplam compare with other treatments for types 1-3 spinal muscular atrophy: a systematic literature review and indirect treatment comparison. *J. Comp. Eff. Res.* 11 (5), 347–370. doi:10.2217/ce-2021-0216
- Roberts, E. K., Elliott, M. R., and Taylor, J. M. G. (2023). Solutions for surrogacy validation with longitudinal outcomes for a gene therapy. *Biometrics* 79 (3), 1840–1852. doi:10.1111/biom.13720
- Russell, S., Bennett, J., Wellman, J. A., Chung, D. C., Yu, Z. F., Tillman, A., et al. (2017). Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* 390 (10097), 849–860. doi:10.1016/S0140-6736(17)31868-8
- Samuel, J. P., Wootton, S. H., and Tyson, J. E. (2023). N-of-1 trials: the epitome of personalized medicine? *J. Clin. Transl. Sci.* 7 (1), e161. doi:10.1017/cts.2023.583
- Serret-Larmande, A., Zenati, B., Dechartres, A., Lambert, J., and Hajage, D. (2023). A methodological review of population-adjusted indirect comparisons reveals inconsistent reporting and suggests publication bias. *J. Clin. Epidemiol.* 163, 1–10. doi:10.1016/j.jclinepi.2023.09.004
- Shah, N., Mojebi, A., Ayers, D., Cope, S., Dhanasiri, S., Davies, F. E., et al. (2022). Indirect treatment comparison of idecabtagene vicleucel versus conventional care in triple-class exposed multiple myeloma. *J. Comp. Eff. Res.* 11 (10), 737–749. doi:10.2217/ce-2022-0045
- Signorovitch, J. E., Sikirica, V., Erder, M. H., Xie, J., Lu, M., Hodgkins, P. S., et al. (2012). Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 15 (6), 940–947. doi:10.1016/j.jval.2012.05.004
- Sola-Morales, O., Volmer, T., and Mantovani, L. (2019). Perspectives to mitigate payer uncertainty in health technology assessment of novel oncology drugs. *J. Mark. Access Health Policy* 7 (1), 1562861. doi:10.1080/20016689.2018.1562861
- Tai, T. A., Latimer, N. R., Benedict, A., Kiss, Z., and Nikolou, A. (2021). Prevalence of immature survival data for anti-cancer drugs presented to the national institute for health and care excellence and impact on decision making. *Value Health* 24 (4), 505–512. doi:10.1016/j.jval.2020.10.016
- Talib, M., and Boon, C. J. F. (2020). Retinal dystrophies and the road to treatment: clinical requirements and considerations. *Asia Pac J. Ophthalmol. (Phila.)* 9 (3), 159–179. doi:10.1097/APO.0000000000000290
- Tanaka, S., Igarashi, A., De Moor, R., Li, N., Hirozane, M., Hong, L. W., et al. (2024). A targeted review of worldwide indirect treatment comparison guidelines and best practices. *Value Health* 27 (9), 1179–1190. doi:10.1016/j.jval.2024.05.015
- Taylor, R. S., and Elston, J. (2009). The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK health technology assessment reports. *Health Technol. Assess.* 13 (8), 1–50. doi:10.3310/hta13080
- Thom, H., Leahy, J., and Jansen, J. P. (2022). Network meta-analysis on disconnected evidence networks when only aggregate data are available: modified methods to include disconnected trials and single-arm studies while minimizing bias. *Med. Decis. Mak.* 42 (7), 906–922. doi:10.1177/0272989X221097081
- Truong, B., Tran, L. T., Le, T. A., Pham, T. T., and Vo, T. T. (2023). Population adjusted-indirect comparisons in health technology assessment: a methodological systematic review. *Res. Synth. Methods* 14 (5), 660–670. doi:10.1002/jrsm.1653
- U.S. Food and Drug Administration, Center for Drug Evaluation and Research (2023). *Considerations for the design and conduct of externally controlled trials for drug and biological products.*
- Velasco Garrido, M., and Mangiapane, S. (2009). Surrogate outcomes in health technology assessment: an international comparison. *Int. J. Technol. Assess. Health Care* 25 (3), 315–322. doi:10.1017/S0266462309990213
- Versteeg, J. W., Vreman, R., Mantel-Teeuwisse, A., and Goetsch, W. (2024). Uncertainty in long-term relative effectiveness of medicines in health technology assessment. *Value Health* 27 (10), 1358–1366. doi:10.1016/j.jval.2024.05.023
- Vestergaard, H. T., D'Apote, L., Schneider, C. K., and Herbets, C. (2013). The evolution of nonclinical regulatory science: advanced therapy medicinal products as a paradigm. *Mol. Ther.* 21 (9), 1644–1648. doi:10.1038/mt.2013.175
- Vinhas, J., Dias, S., Gouveia, A. M., Correia, A., Dias, C. V., Sousa, D., et al. (2021). “Pharmacotherapeutic assessment methodology, version 3.0,” in Committee for Health Technology Assessment, INFARMED – Autoridade Nacional do Medicamento e Produtos de Saúde.
- Vreman, R. A., Broekhoff, T. F., Leufkens, H. G., Mantel-Teeuwisse, A. K., and Goetsch, W. G. (2020). Application of managed entry agreements for innovative therapies in different settings and combinations: a feasibility analysis. *Int. J. Environ. Res. Public Health* 17 (22), 8309. doi:10.3390/ijerph17228309
- Wheaton, L., and Bujkiewicz, S. (2025). Use of surrogate endpoints in health technology assessment: a review of selected NICE technology appraisals in oncology. *Int. J. Technol. Assess. Health Care* 41 (1), e11. doi:10.1017/S0266462325000017
- Wheaton, L., Papanikos, A., Thomas, A., and Bujkiewicz, S. (2023). Using bayesian evidence synthesis methods to incorporate real-world evidence in surrogate endpoint evaluation. *Med. Decis. Mak.* 43 (5), 539–552. doi:10.1177/0272989X231162852
- Woodcock, J., and LaVange, L. M. (2017). Master protocols to study multiple therapies, multiple diseases, or both. *N. Engl. J. Med.* 377 (1), 62–70. doi:10.1056/NEJMra1510062
- Zorginstituut Nederland (2024). *Guideline for economic evaluations in healthcare.*