BMJ Open

Which clinical trial designs and statistical approaches have been used in assessments of orphan maintenance by the European Medicines Agency between 2012 and 2022? A cross-sectional study.

Fabian Windfuhr ,[1,2] Kristina Larsson,[1] Theodor Framke,[1,3] Florian Lasch[1,3]

To cite: Windfuhr F, Larsson K, Framke T, *et al*. Which clinical trial designs and statistical approaches have been used in assessments of orphan maintenance by the European Medicines Agency between 2012 and 2022? A cross-sectional study.. *BMJ Open* 2024;**14**:e086171. doi:10.1136/bmjopen-2024-086171

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-086171).

Received 11 March 2024
Accepted 25 October 2024

[1]European Medicines Agency, Amsterdam, The Netherlands
[2]Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[3]Institute for Biostatistics, Hannover Medical School, Hannover, Germany

Correspondence to
Dr Florian Lasch;
florian.lasch@ema.europa.eu

ABSTRACT

Objectives In the European Union, a new orphan medicinal product must demonstrate 'significant benefit' over approved medicinal products targeting the same indication. To demonstrate a significant benefit, comparisons between the new product and the already approved medicinal products—either directly by a head-to-head comparison within a clinical trial or indirectly as a cross-trial comparison—are necessary. In this study, we investigate the types of trial designs and statistical approaches used for demonstrating a significant benefit of a new orphan medicinal product against approved comparators used between 2012 and 2022.

Design This is a cross-sectional study based on the European Medicines Agency's 'orphan maintenance' assessment documents between 2012 and 2022. All documents were manually reviewed to extract structured data on the following outcome measures:
For every comparison between a new orphan medicinal product and a comparator used for demonstrating a significant benefit as part of an orphan maintenance procedure, we recorded the type and design of the data source and the type of statistical methodology used for the comparison.

Results We identified 151 EMA orphan maintenance procedures with a positive decision that required the demonstration of a significant benefit. Within these 151 procedures, 418 comparisons between medicinal products were identified. Indirect comparisons are the most common approach for comparing the new orphan medicinal product to a relevant comparator (44%, 182/418), followed by qualitative comparisons (39%, 162/418) and direct comparisons (18%, 74/418). Among the indirect comparisons, naive side-by-side comparisons are most often used (71%, 129/182), whereas inferential approaches that adjust for population differences and quantify the uncertainty of the comparison are used less often (29%, 53/182). Although there is no clear time trend in the prevalence of any specific comparison type, we find that inferential indirect comparison methods approximately doubled between the first and second half of the reviewed time frame.

Conclusions Indirect comparisons play an important role in demonstrating a significant benefit in the assessment of orphan products. Further work is needed to evaluate the appropriateness of different methodologies.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Strength: This review is not based on a random sample but includes all EMA orphan maintenance procedures with a positive outcome between 2012 and 2022.
⇒ Strength: Access to all submitted documentation from applicants allowed a precise evaluation and categorisation of the proposed data and methods.
⇒ Limitation: This review focused on EMA orphan maintenance procedures with a positive outcome, since applicants mostly withdraw applications before a negative outcome is concluded and hence final data on methods and their evaluation is lacking.

INTRODUCTION

In 2000, the Regulation (EC) No. 141/2000 on orphan medicinal products became effective in the European Union (EU). The legislation was introduced to incentivise the development of medicinal products in populations affected by rare diseases. More than 20 years down the line, there is clear evidence that the EU Orphan Regulation has made important contributions to the overall development of new medicines for rare diseases, both by improving the environment for research and development, and by providing economic incentives to developers. The regulation and the general focus on rare diseases have brought benefits to patients.[1]

In the EU, rarity is defined as a condition not affecting more than 5 in 10 000 persons.

An additional requirement is that if 'satisfactory methods' to treat the condition are approved, the medicinal product applied for must be of 'significant benefit' to those affected by that condition.[2] Any medicinal product approved in the EU for the condition is generally considered a satisfactory treatment method.

Significant benefit can be defined either as a clinically relevant advantage or a major contribution to patient care. It is assessed in comparison with all products approved for the therapeutic indication at the time of both initial orphan designation and marketing authorisation of an orphan medicinal product. When a pharmaceutical company seeks an orphan designation, it is usually given at an early time point in the development of the medicinal product; therefore, only very limited data will be available, and the assumed significant benefit is often uncertain. At the time of the marketing authorisation, it has to be assessed whether the orphan criteria are still met, that is, 'maintained', hence it is called the orphan maintenance procedure. The Committee for Orphan Medicinal Products (COMP) is the central body responsible for evaluating applications for (maintenance of) orphan designation. It consists of one expert from each EU and European Economic Area (EEA) member state, as well as three patient representatives and additional topic experts. The COMP is responsible for evaluating whether applications fulfil the regulatory requirements for orphan designation, such as significant benefit.

Criteria for demonstration of a significant benefit

Based on the Commission notice on the application of Articles 3, 5 and 7 of Regulation (EC) No. 141/2000 on orphan medicinal products (2016/C 424/03), one of the following two criteria needs to be fulfilled:

- A 'clinically relevant advantage' may be based on:
  - improved efficacy for the entire population suffering from the condition or a particular population subset or a subset that is resistant to the existing treatments, or
  - a better safety profile or a better tolerability for the entire population suffering from the condition or for a particular subset.
- A 'major contribution to patient care' may be based on:
  - ease of self-administration, for example, if the new treatment allows ambulatory treatment instead of treatment in a hospital only or if it has a significant impact on convenience of use and reduces treatment burden, or
  - significantly improved adherence to treatment due to a change in pharmaceutical form (eg, modified release formulation), provided there are documented difficulties with the existing form and data showing better clinical outcomes with the new form.

Drug development in rare conditions faces many challenges. In particular, difficulties are encountered in conducting well-powered clinical trials due to the limited patient population. Even though there is guidance on how to design and optimally use data from trials in rare disorders,[3 4] the issue remains that the development of medicinal products in a small population is challenging and the same robustness as can be expected from trials in non-rare diseases might not be feasible.[5] In principle, randomised controlled trials (RCTs) of the candidate orphan medicinal product against all other available satisfactory methods would provide the highest quality evidence for establishing a significant benefit. However, the rarity and heterogeneity of conditions and the complexity of the treatment algorithms complicate the demonstration of significant benefit via one or multiple RCTs.

Therefore, alternative methods like indirect comparisons of the new treatment against comparator products may be used to establish the significant benefit of the new treatment over the existing comparator products.[6]

Indirect treatment comparisons (here abbreviated as IC; in the literature also occasionally abbreviated as ITC) allow the cross-trial comparison of interventions that have not been directly compared in the same clinical trial. Fundamentally, an indirect comparison is based on data from two or more different trials. Importantly, in this situation, the trials may have included different patient populations. Various methods exist to compare the effects observed in different trials. To overcome the main limitation of data from different trials not being comparable, various methodological approaches have been developed for adjusting observed population differences (eg, different distributions of demographic characteristics). The available methods for indirect comparisons include simple (unadjusted) methods like the side-by-side (SBS) comparison, over-adjusted methods like the matching-adjusted indirect comparison (MAIC)[7] and more complex approaches taking into account whole networks of evidence of available treatments in a given indication (eg, network meta-analysis [NMA]).[8] Methodological approaches have been developed to use only aggregate data, a mix of aggregate data and individual patient data (IPD) or only IPD.[9] In this context, the possibility of assessing or adjusting for the difference between populations is further determined based on the reports of the different trials (the set of baseline variables reported and whether the trial sponsor makes IPD available for patient-level analyses).

Anecdotal evidence and findings from a recent report suggested that indirect comparisons have been used more in recent years in support of the significant benefit at the time of marketing authorisation for orphan medicinal products, and that more sophisticated methodologies like NMAs and MAICs were .[10]

To investigate the hypothesis that indirect comparison methods are increasingly used for demonstrating a significant benefit, we conducted a systematic evaluation of the role of indirect comparisons in the context of demonstrating significant benefit for orphan medicines as part of the orphan maintenance decision at the time

of marketing authorisation assessment, addressing the following questions:

1. How many orphan maintenance procedures with a positive opinion use indirect comparison methodology?
2. Which statistical methods are proposed by applicants and accepted by the COMP for indirect comparisons?
3. Are there differences between therapeutic areas?
4. Is there a trend over time?

To investigate these questions and to derive a complete picture of the methodologies used for indirect comparison, we conducted a review of EMA COMP procedures with positive outcomes in the past 11 years following the methodology described in the Methods section.

METHODS

Study design and selection of EMA orphan maintenance procedures

We performed a retrospective cohort study of EMA maintenance of orphan designation procedures between 2012 and 2022 in which significant benefit had to be demonstrated. This scope ensured that all included orphan maintenance procedures contained a direct or indirect comparison against competitors on the market. To obtain an overview of the currently accepted practice in efficacy comparisons as part of demonstrating significant benefit, we only included orphan maintenance procedures from 2012 to 2022 with a positive outcome in our review. More concretely, all orphan maintenance procedures pertaining to products with a marketing authorisation date (thus given a positive opinion by the Committee for Human Medicinal Products (CHMP), hereafter the date of the positive opinion is termed 'birth date') between 01 January 2012 and 31 December 2022 were selected from EMA's internal database of documents. In our subsequent time-dependent analyses, however, the date of the COMP decision was used as it better reflects the timing of the COMP evaluation of each procedure. Therefore, two orphan maintenance procedures date back to 2011 in the data set, which are visible in all plots displaying time as a variable.

Orphan maintenance procedures were included, irrespective of procedure type (initial marketing authorisations or extensions of indication) and also disregarding whether the orphan status was later withdrawn or whether their marketing exclusivity expired during the study period. All satisfactory methods reflect the state at the time of the report irrespective of later decisions (ie, the outcome of a court case). The review of the methodology used for demonstrating significant benefit was based on the applicant's submission documents and the scientific assessment report compiled by the COMP. These COMP reports (published on the EMA webpage as Orphan Maintenance Assessment Report since 2018), are a summary of the sponsor-supplied data, as well as the assessment of the data and regulatory considerations by the committee. If the COMP issued a list of questions on the significant benefit, this document and the applicant's response were also reviewed and any additional relevant comparisons were included in the review.

DATA COLLECTION

Each orphan maintenance procedure may include several comparisons; therefore, information on two levels needs to be considered—the procedure level and the comparison level. All documents were manually reviewed to extract the following information:

On the procedure level, we recorded:

- the name of the product under review
- the indication of the product under review
- the COMP's opinion
- the grounds for this opinion
- the number of comparators, defined as any product identified as a satisfactory method of the respective procedure. Importantly, when a product was compared against the standard of care or best available therapy, 'best available therapy' or 'standard of care' was considered as one comparator.
- whether a list of questions regarding the product's significant benefit was issued or not.

For each of the comparisons, defined as a comparison of the product under review against a satisfactory method (as identified in the report section 'Criteria for demonstration of a significant benefit'), we recorded:

- information on the comparison method and categorised the type of comparison methods (see table 1 for categories).
- the design of the trial of the orphan drug
- the design of the trial/data source of the comparator
- the COMP's appraisal of each comparison.

Importantly, because of this data structure, some analyses presented in the Results section represent frequencies relative to the absolute number of orphan maintenance procedures, whereas most analyses display frequencies related to the absolute number of comparisons. Details on the definition of the comparison, trial designs and appraisal outcome can be found in online supplemental material 1.

Patient and public involvement

Patients or the public were not involved in the design or conduct of this study. However, the study results were presented to the COMP, which includes patient representatives, and all feedback data received through this process were incorporated into the manuscript.

Statistical analysis

The data management and statistical analysis of all collected information was performed with the R software,[11] using the readxl, lubridate, tidyverse, ggplot2, scales and reshape2 packages. The main aim of the data analysis was to quantify the absolute and relative frequency of the use of different comparison methods,

Table 1 Occurrence and a short description of all comparison methods identified in the reviewed sample; the chosen categorisation into five larger categories is reflected in all figures describing the identified comparisons and was chosen to reflect the most important methodological differences between the comparisons

Category	Method	n=418	Short description
Quantitative, direct comparisons 74 (18%)			
	Head-to-head comparison	60	Direct comparison of two products as two parallel arms of one study, such as in a randomised controlled trial
	Baseline comparison	14	Comparing the outcome of one product measured at the baseline of a study and the outcome of the other product at the end of the study
Quantitative, indirect comparisons 182 (44%)			
Side-by-side comparisons (n=129, 31%)	Simple side-by-side comparison	113	Presentation of summary statistics for a variable (eg, objective response rate for 'response') by treatment arms. The treatment arms are from separate studies, and no statistical methods for cross-trial comparisons are applied (eg, the difference between objective response rates from different studies)
	Pooled side-by-side comparison	16	Same as the simple side-by-side comparison, but the effect size from one or more of the comparators is derived from pooling results from several studies
Inferential comparison with aggregate external data (n=40, 10%)	Matching-adjusted indirect comparison	22	Comparing individual patient data from the investigational product with aggregate data from one comparator from another study by means of re-weighting the individual patient data to match the baseline characteristics of the aggregate comparator data[7]
	Simulated treatment comparison	8	A regression-based approach estimating the effect of an investigational product based on individual patient data and adjusted for baseline characteristics compared with aggregate data for the comparator. The approach can have the additional element of simulation where samples are drawn from the joint covariate distribution of the aggregate data)[18]
	Bucher method	7	Compares two or more products that have the same comparator (eg, placebo) via indirect adjustment[19]
	Meta-analysis	1	Estimates the effects of two products using aggregate data from at least two independent studies. The combined (pooled) effect estimate is based on the weighted average of the independent studies[20]
	Network meta-analysis	2	Compares more than two products with data from independent studies by combining direct and indirect evidence, here based on aggregate data[21]
Inferential comparison with patient-level external data (n=13, 3%)	Matched/weighted comparison	4	Indirect comparison based on matching patient-level data from each patient under the investigational treatment to data from the control group, or weighting data from the control group depending on their similarity to the treated patients (often weighted by the probability to receive the treatment based on several variables measured in treated and untreated patients) to create a comparable control group
	Regression	4	Compares two products based on patient-level data in a regression model (eg, linear regression or Cox regression)
	Network meta-analysis	5	Compares more than two products with data from independent studies by combining direct and indirect evidence, here based on individual patient data[21]
Qualitative comparison		162 (39%)	
	Partial overlap in the patient population	50	Instances where there was no complete overlap in indications for two products
	Non-preferred treatment	44	Any products marketed as non-preferred treatments, for example, second- or later-line products, therefore not needing to show improvement over earlier line/preferred products
	Adjunct treatment	47	Instances in which the investigational product is supposed to be used in combination with the comparator
	Unclear	21	All those instances, in which no quantitative comparison could be clearly identified

both by combining the overall time frame and by year to investigate time trends. The overall approach to the analyses is descriptive; no inferential methods were applied.

RESULTS

General characteristics of the selected EMA orphan maintenance procedures

Overall, 151 orphan maintenance procedures were identified matching the inclusion and exclusion criteria. Within the specified time frame, this was a subset of around 52% (151/297) of all orphan maintenance procedures (irrespective of outcome), and around 78% (151/197) of all orphan maintenance procedures that received a positive opinion, regardless of whether significant benefit had to be demonstrated or not (see online supplemental material 2 for more details). Across these 151 orphan maintenance procedures, between 1 and 10 comparators per procedure (median=3, IQR=2–4; see figure 1) were noted.

In approximately half of all cases, a list of questions was issued regarding the significant benefit. The final

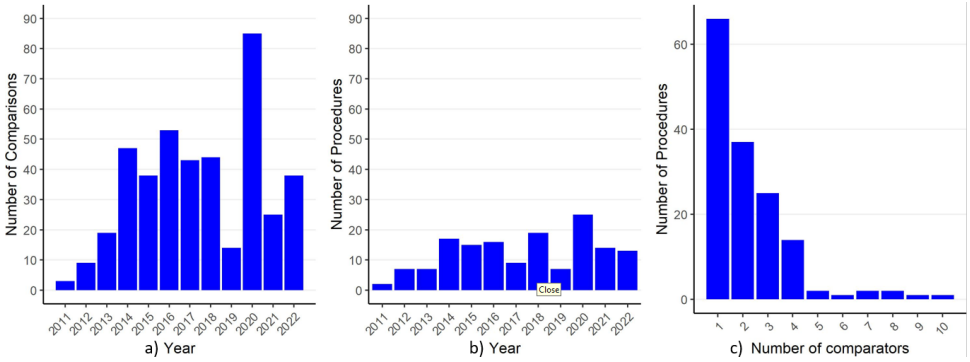

Figure 1 (a) Absolute frequency of comparisons, (b) orphan maintenance procedures and (c) comparators per procedure.

positive opinion was based on a clinically relevant advantage in the majority of orphan maintenance procedures, but there were also several orphan maintenance procedures based on a major contribution to patient care, as well as on a combination of a clinically relevant advantage and a major contribution to patient care (see online supplemental material 2 for an overview).

Using the system organ classes by the medical dictionary for regulatory activities categories[12] for categorising the disease areas, 40% (60/151) of the orphan maintenance procedures concerned 'Blood and lymphatic system disorders', making it the most targeted disease area in the sample. Products for indications, such as multiple myeloma and diffuse large B-cell lymphoma, would be found in this category. This was followed by 'Congenital familial and genetic disorders' with 19% (28/151) of orphan maintenance procedures and 'Neoplasms benign, malignant and unspecified' with 12% (18/151) of orphan maintenance procedures, where, for example, cystic fibrosis and ovarian cancer would be included. Any other MedDRA categories were subject to eight or less orphan maintenance procedures (for an overview, see online supplemental material 3). More broadly, 45% (68/151) of the orphan maintenance procedures were based on an oncological indication.

Overall, 418 comparisons were identified across all the 151 orphan maintenance procedures (median=2, IQR=1–3, range=1–14). Sixteen different types of comparison methods were identified, which were categorised into five broader groups of comparison types (see table 1).

Regarding the trial designs of the data sources underlying these comparisons, RCTs represented the majority of cases with 68% (284/418) of all main trials and 49% (206/418) of all comparator trials. single-arm trials (SATs) were the next most frequent type of trial design and were used as a source in 28% (116/418) of all main trials and 7% (28/418) of all comparator trials. For a full overview of trial designs, see online supplemental material 4.

Frequency of different comparison methods and development over time

Indirect comparisons are the most common approach for comparing the new orphan medicinal product to a relevant comparator (44%, 182/418), followed by qualitative comparisons (39%, 162/418) and direct comparisons (18%, 74/418) (see figure 2 and table 1). Among the indirect comparisons, naive SBS comparisons are most often used (71%, 129/182) whereas inferential approaches that adjust for population differences or quantify the uncertainty of the comparison, either using or not using IPD, are less often used (29%, 53/182).

Comparing the first and second half of the investigated time frame, between 2011 and the end of 2017, 6% (12/212) of the identified comparisons were based on inferential methods (regardless of the use of IPD), whereas from January 2018 until December 2022, 20% (41/206) of the comparisons were based on inferential methods. When looking at SBS comparisons, 37% (79/212) were identified in the first half and 24% (50/206) were identified in the second half of the reviewed time frame (for an overview, see figure 2c). Therefore, while the relative frequency of the other types of quantitative comparisons declined slightly, the proportion of inferential indirect comparison methods approximately doubled between the first and second half of the reviewed time frame.

Acceptance of different comparison methods by the COMP

Generally, the acceptability of a comparison by the COMP depends on both the comparison method and the data. If a comparison was accepted, the comparison method was accepted in the specific situation.

The comparison method with the highest relative frequency of acceptance was qualitative comparisons followed by direct comparisons. Conversely, the proportion of rejected comparisons was highest among the indirect comparisons, specifically the inferential methods using aggregate data. However, most rejections specifically based on the methodological limitations of the comparison type were observed for the SBS comparisons (for an overview, see figure 3; for more details, see online supplemental file 5).

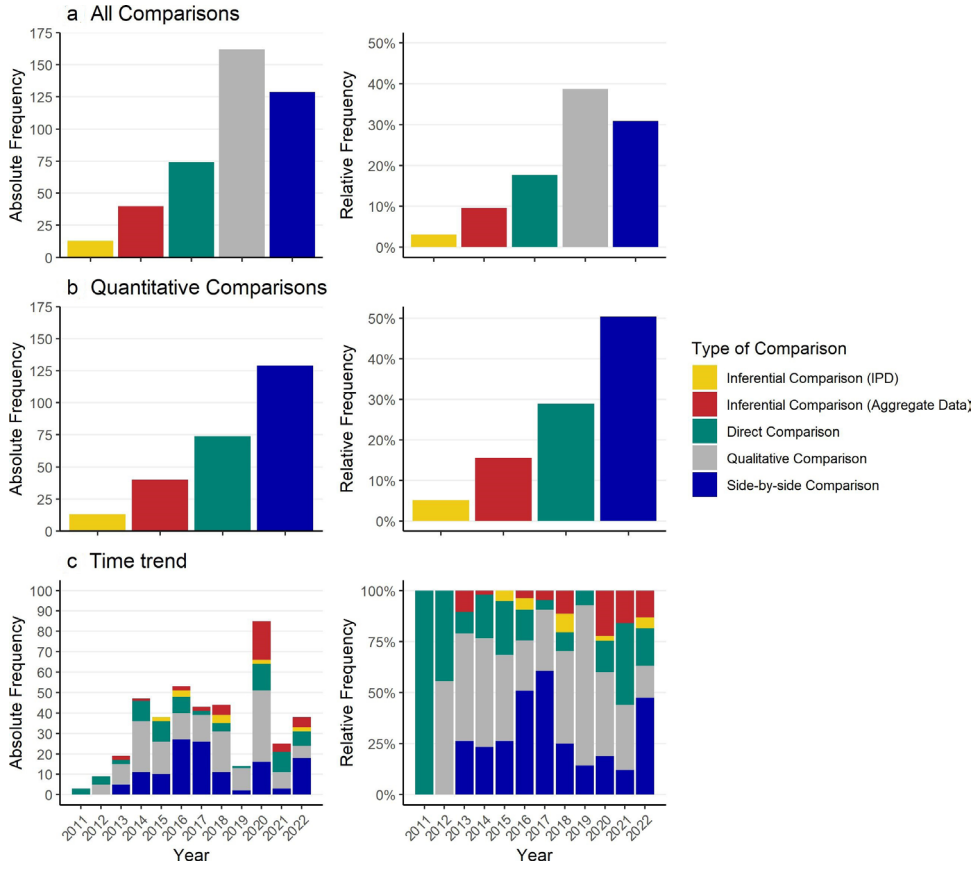

Figure 2 Absolute (left) and relative frequency (right) of different types of comparisons: (a) all identified comparisons, (b) quantitative comparisons only and (c) distribution of comparison types per year. IPD, individual patient data.

To explore the appraisal of the different comparison methods by the COMP further, we also analysed the number of cases in which the COMP raised a list of questions regarding the significant benefit. A list of questions is issued if COMP has remaining questions concerning the comparisons that are proposed by the applicant. Following the list of questions, the applicant prepares a response to these questions for evaluation, most often with new methods applied to the same data. We found a higher proportion of indirect comparisons and a lower proportion of direct and qualitative comparisons in orphan maintenance procedures with a list of questions (see figure 4, top two panels (a)).

Differences between therapeutic areas

To investigate potential differences between therapeutic areas regarding the choice of comparison methods, we

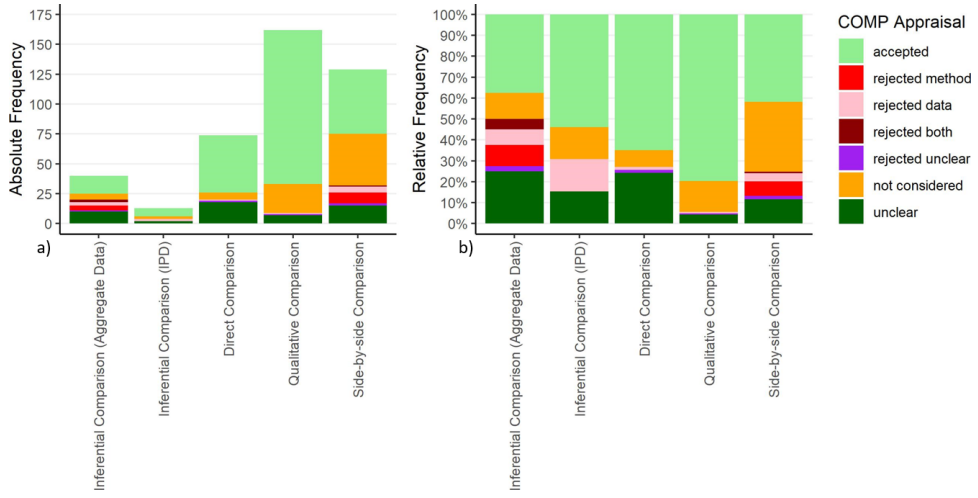

Figure 3 Absolute (a) and relative frequency (b) of the COMP's appraisal of comparisons. COMP, Committee for Orphan Medicinal Products; IPD, individual patient data.

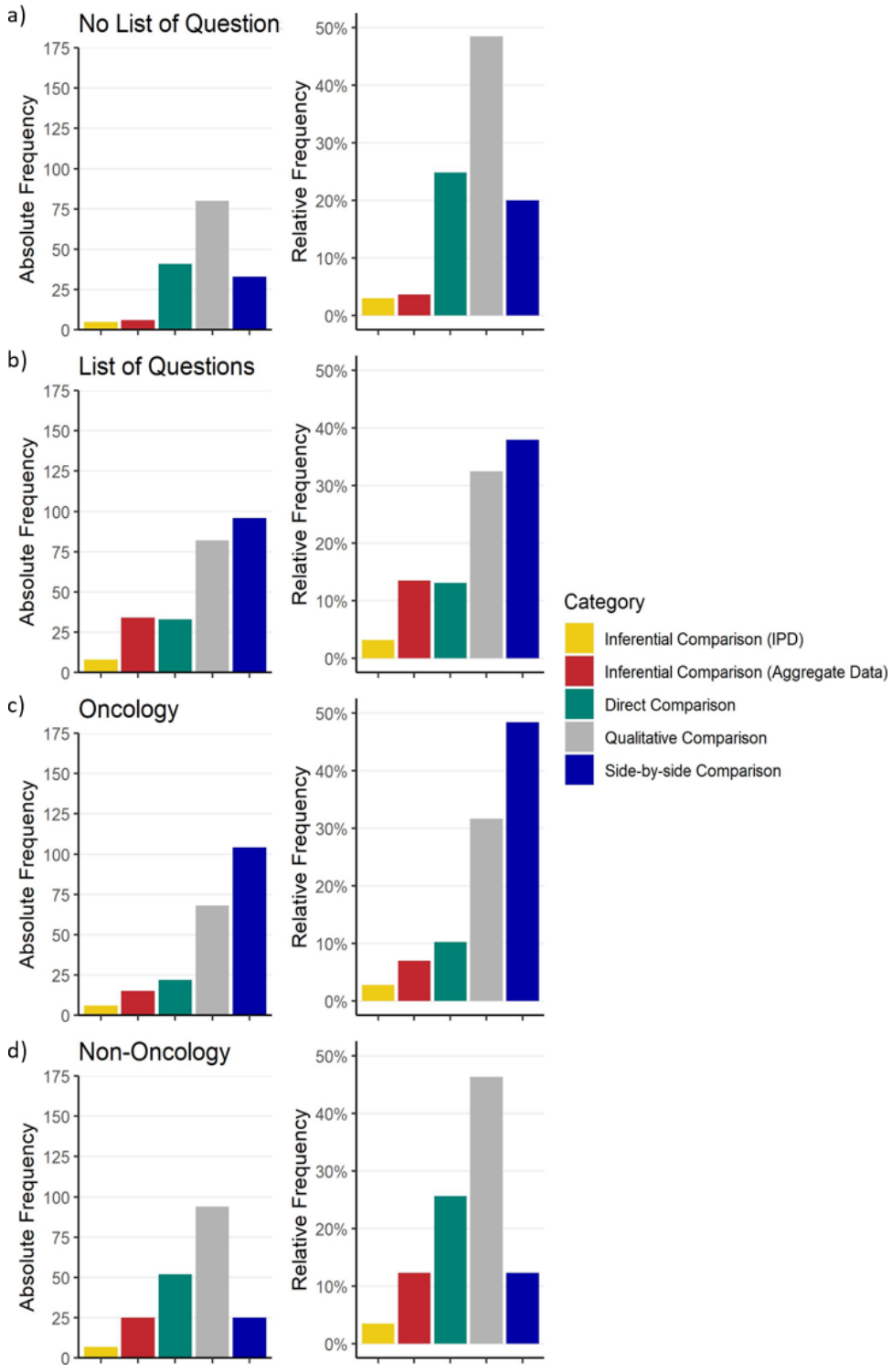

Figure 4 Absolute (left) and relative frequency (right) of different types of comparisons across two stratifications. a) Orphan maintenance procedures in which a list of questions regarding the significant benefit was not issued and b) procedures were a list of questions was issued. c) All oncology procedures compared with d) all non-oncology orphan maintenance procedures. IPD, individual patient data.

distinguished all reviewed orphan maintenance procedures into oncology and non-oncology orphan maintenance procedures. While other categorisations would have been interesting to investigate, the distribution of therapeutic areas and the high proportion of oncology did not allow meaningful comparisons within the non-oncology indications.

Non-oncology orphan maintenance procedures were supported by direct comparisons 2.5 times more often than oncology orphan maintenance procedures, that

is, ca. 25% of cases in non-oncology against 10% within oncology. Investigating the indirect comparison methods used, in oncology, 50% of the comparisons were SBS comparisons. In contrast, SBS comparisons made up a little over 10% of non-oncology orphan maintenance procedures. The use of inferential indirect comparison methods, however, was higher in non-oncology orphan maintenance procedures (for an overview, see figure 4).

Further differences between oncology and non-oncology orphan maintenance procedures can be seen regarding the trial design and appraisal of the comparison method. SATs were the basis for comparisons far more often in oncology orphan maintenance procedures than in non-oncology orphan maintenance procedures (32%, 68/215 SATs for the pivotal trial design in oncology vs 15%, 30/203 in non-oncology; see online supplemental files 6 and 7). Yet, RCTs were still the most used data source for pivotal trials as well as comparator trials, in both non-oncology and oncology orphan maintenance procedures. Looking at the COMP's appraisal, our data show that a lower proportion of comparisons was rejected in oncology orphan maintenance procedures, particularly among all indirect comparisons (see online supplemental file 8).

DISCUSSION

This review of orphan maintenance procedures of the EMA COMP has investigated how a significant benefit has been demonstrated by applicants. Furthermore, for the cases where an indirect comparison between the new product and already licensed products was performed, we have explored the types of approaches that have been used.

Overall, a high number of qualitative comparisons were used to demonstrate significant benefit. The reason for this observation is the definition of a 'satisfactory method' in the orphan regulation, determining the necessary comparators against which to demonstrate a significant benefit. Since a satisfactory method must be approved for an overlapping therapeutic indication, in case of partial overlaps between the indications of the comparator and the new product, the significant benefit can be based on these additional patients who cannot be treated with the approved products. In the oncology setting, the main drivers of the qualitative assessment are the approvals in the (last-line) setting where no other products are approved, and the patients have been treated with the approved products in earlier lines of treatment. On the contrary, in the non-oncology setting, the qualitative comparisons are not driven by treatment lines but by a partial or no overlap of indications and adjunctive treatments.

Additionally, we have observed a wide span in the number of comparators, ranging from 1 to 10 comparators per product, which likely reflects the diverse situation across therapeutic areas and corresponding variability in the number of products approved per condition. For example, in multiple myeloma and cystic fibrosis, there are numerous medicinal products approved to treat different aspects and stages of the disease whereas, for other conditions like cystinosis and myasthenia gravis, only very few medicinal products are approved at the time of assessment of a new treatment.

Comparing the type of indirect comparison methods between oncology and non-oncology indications shows a notable difference in comparison methods and COMP appraisal that requires further investigation. While in oncology, SBS comparisons are the most used method, for non-oncology products, qualitative comparisons followed by direct comparisons were most prominent. Correspondingly, a previous study found that around one-quarter of all pivotal trials used in EMA approvals of oncology products 2014–2016 were single-arm trials.[13] According to our data, the proportion of rejected comparisons was lower in oncology compared with non-oncology indications. For context, prior research investigating the difference in overall approval rates between oncology and non-oncology products found that, in EMA procedures between 2009 and 2018, oncology products were approved marginally less often than non-oncology products.[14] Meanwhile, it has also been reported that oncology products approved by EMA often provide little or no added benefits,[15] though no distinction between orphan or non-orphan products has been made in the analysis. In the context of orphan medicinal products, more research is needed to elucidate whether there might be different evidentiary standards across indications, or if there are any differences in the actual added benefits the products bring.

In the present study, looking at the overall sample regardless of indication, we also found that more than 25% of the quantitative comparisons were direct comparisons. This observation highlights that the rarity of a disease per se does not prohibit or prevent the conduct of RCTs.

Evaluating the COMP's appraisal of different comparison methods shows that qualitative comparisons and direct comparisons were accepted in most cases, whereas indirect comparisons were accepted less often. SBS comparisons were accepted less often as an indirect comparison method than approaches that adjust for differences between populations. While the hypothesised overall increase in indirect comparisons could not be found in the available data, the increase in indirect comparisons using more sophisticated statistical methods was partly confirmed. Even though the yearly analysis did not show a continuous increase between 2011 and 2022, we have seen that the proportion of indirect comparisons using inferential statistical methods nearly doubled from 2011–2017 to 2017–2022. Considering that over time more products have been approved for many rare diseases and the continued developments in NMA techniques, the importance of inferential statistical methods for indirect comparisons might further increase in the future. Also considering the challenges of Health Technology

Assessment after new medicines have been licensed, our findings highlight the need for adequate planning of clinical trials that need to meet the requirements of different decision-makers. The need to conduct indirect comparisons should be anticipated at the trial design stage with a view on how the new trial fits into the evidence network to ensure that the necessary variables for using statistical methods for indirect comparisons that adjust for differences between populations are collected in the trial. On a general note, in many instances, only aggregate data were available for the comparator against which the new orphan medicine needed to be compared. If marketing authorisation holders would make their data readily available, this could increase the quality of the indirect comparisons as it would enable the use of better statistical methodologies, ultimately facilitating better decisions in the interest of patients.

For medicinal product licensing in the EU, indirect comparisons are not only relevant for demonstrating a significant benefit as part of the orphan maintenance procedure. In the context of conditional marketing authorisation through the EMA CHMP, indirect comparisons can also play a role in demonstrating a major therapeutic advantage. After drug licensing, indirect comparisons play a crucial role in determining the relative effectiveness of authorised treatments as part of the health technology assessment. It would be interesting to explore similarities and differences in the use of indirect comparison approaches between these different fields of application.

In conclusion, indirect comparisons already are and will continue to be an important tool in the assessment of orphan products' significant benefit at the time of marketing authorisation. While health technology assessment bodies regularly use and provide guidance on indirect comparison methods to compare the relative effectiveness of a new medicinal product,[16 17] further work is needed to understand the appropriateness of indirect comparison approaches for demonstrating a significant benefit, guiding the sponsor's choices and the regulatory assessment.

In this review, we have only included orphan maintenance procedures with a positive outcome. This choice was mainly driven by considerations of data accessibility. Most non-positive COMP opinions result in the applicant removing the orphan status voluntarily and progressing with a non-orphan marketing authorisation. Therefore, these assessments do not reach a conclusion and, in many cases, no final COMP opinion would have been documented describing the acceptability of the indirect comparison methodologies. In addition, in our review period, only eight orphan maintenance procedures resulted in a negative opinion, which was considered too small for meaningful comparisons.

We focused on orphan maintenance decisions of the COMP; however, indirect comparisons can also play a role in the initial orphan designations. To derive a complete picture of the use of indirect comparison for COMP decisions, it would be interesting to expand this review to orphan designation decisions in the future. The limited number of orphan maintenance procedures prevented the investigation of multiple factors at the same time (e.g., rarity of the disease, comparison type and COMP appraisal).

In conclusion, indirect comparisons already are and will continue to be an important tool in the assessment of orphan products' significant benefit at the time of marketing authorisation. While health technology assessment bodies regularly use and provide guidance on indirect comparison methods to compare the relative effectiveness of a new medicinal product,[16 17] further work is needed to understand the appropriateness of indirect comparison approaches for demonstrating a significant benefit, guiding the sponsor's choices and the regulatory assessment.

Acknowledgements We thank Dr. Frauke Naumann-Winter, Dr. David King, Dr. Celia Castaño-Amores and Dr. Devidas Menon for helpful comments.

Contributors All authors were involved in the planning of this study. FW carried out the data collection and data analysis. KL, TF and FL supervised the project. All authors collaboratively discussed the results, contributed to the writing of the manuscript and are guarantors.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer The views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iD

Fabian Windfuhr http://orcid.org/0009-0003-6065-2860

REFERENCES

1 European Commission. Evaluation of the medicines for rare diseases and children legislation. 2020. Available: https://health.ec.europa.eu/medicinal-products/medicines-children/evaluation-medicines-rare-diseases-and-children-legislation_en

2 European Commission. Regulation (EC) no 141/2000 of the European parliament and of the council of 16 December 1999 on orphan medicinal products. 1999. Available: http://data.europa.eu/eli/reg/2000/141/oj
3 Day S, Jonker AH, Lau LPL, *et al*. Recommendations for the design of small population clinical trials. *Orphanet J Rare Dis* 2018;13:195.
4 European Medicines Agency Committee for Human Medicinal Products. Guideline on clinical trials in small populations. 2006. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-trials-small-populations_en.pdf
5 Pontes C, Fontanet JM, Vives R, *et al*. Evidence supporting regulatory-decision making on orphan medicinal products authorisation in Europe: methodological uncertainties. *Orphanet J Rare Dis* 2018;13:206.
6 European Commission. Commission notice on the application of articles 3, 5and 7 of regulation (EC) no 141/2000 on orphan medicinal products. 2016 Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:JOC_2016_424_R_0003&from=EN
7 Signorovitch JE, Sikirica V, Erder MH, *et al*. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 2012;15:940–7.
8 Dias S, Sutton AJ, Ades AE, *et al*. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making* 2013;33:607–17.
9 Veroniki AA, Straus SE, Soobiah C, *et al*. A scoping review of indirect comparison methods and applications using individual patient data. *BMC Med Res Methodol* 2016;16:47.
10 Naumann-Winter F, Wolter F, Hermes U, *et al*. Licensing of Orphan Medicinal Products-Use of Real-World Data and Other External Data on Efficacy Aspects in Marketing Authorization Applications Concluded at the European Medicines Agency Between 2019 and 2021. *Front Pharmacol* 2022;13:920336.
11 Core Team RC. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna; 2013.
12 Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf* 1999;20:109–17.
13 Naci H, Davis C, Savović J, *et al*. Design characteristics, risk of bias, and reporting of randomised controlled trials supporting approvals of cancer drugs by European Medicines Agency, 2014-16: cross sectional analysis. *BMJ* 2019;366:l5221.
14 Rohr U, Iovino M, Rudofsky L, *et al*. A decade comparison of regulatory decision patterns for oncology products to all other non-oncology products among Swissmedic, European Medicines Agency, and US Food and Drug Administration. *Clinical Translational Sci* 2023;16:1569–81.
15 Brinkhuis F, Goettsch WG, Mantel-Teeuwisse AK, *et al*. Added benefit and revenues of oncology drugs approved by the European Medicines Agency between 1995 and 2020: retrospective cohort study. *BMJ* 2024;384:e077391.
16 Phillippo DM, Ades AE, Dias S, *et al*. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Med Decis Making* 2018;38:200–11.
17 EUnetHTA-21. Individual practical guideline document D4.3.1: direct and indirect comparisons. 2022. Available: https://www.eunethta.eu/wp-content/uploads/2022/12/EUnetHTA-21-D4.3.1-Direct-and-indirect-comparisons-v1.0.pdf
18 Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. *Pharmacoeconomics* 2015;33:537–49.
19 Bucher HC, Guyatt GH, Griffith LE, *et al*. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683–91.
20 Deeks JJ, Higgins JP, Altman DG. Cochrane statistical methods group. analysing data and undertaking meta-analyses. In: *Cochrane handbook for systematic reviews of interventions*. 2019: 241–84.
21 Chaimani A, Caldwell DM, LI T, *et al*. Undertaking network meta-analyses. In: *Cochrane handbook for systematic reviews of interventions*. 2019: 285–320.