BMJ Open

Clinical Trial designs and statistical approaches for indirect comparisons in the assessment of the Orphan Designation status: An overview from 2012 to 2022

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-086171
Article Type:	Original research
Date Submitted by the Author:	11-Mar-2024
Complete List of Authors:	Windfuhr, Fabian; European Medicines Agency, ; University Medical Centre Groningen Larsson, Kristina; European Medicines Agency Framke, Theodor; European Medicines Agency; Institute for Biostatistics Lasch, Florian ; European Medicines Agency; Institute for Biostatistics
Keywords:	Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Public health < INFECTIOUS DISEASES, Research Design

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Clinical Trial designs and statistical approaches for indirect comparisons in the assessment of the Orphan Designation status: An overview from 2012 to 2022

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

[1] European Medicines Agency, Domenico Scarlattilaan 6, 1083 HS Amsterdam, The Netherlands
[2] University Medical Centre Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands
[3] Hannover Medical School, Institute for Biostatistics, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

Key words: *Orphan designation, significant benefit, indirect comparison, indirect treatment comparison, evidence synthesis.*

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.

ABSTRACT

In the European Union, a new orphan medicinal product must demonstrate 'significant benefit' over approved medicinal products targeting the same indication, then defined as "satisfactory methods". To demonstrate a significant benefit, comparisons between the new product and the all satisfactory methods – either directly by a head-to-head comparison within a clinical trial, or indirectly as a cross-trial comparison - are necessary. However, the role of indirect comparisons, the frequency of their use and the type of indirect comparisons methods used have not been systematically assessed. Therefore, in the present study, we reviewed all maintenance of orphan designation procedures with a positive outcome in which significant benefit had to be demonstrated. Overall, we find that indirect comparisons make up most proposed comparisons. Within these, the most prevalent type are naive comparisons. Although there is no clear trend in the prevalence of any specific comparison type, we find that inferential indirect comparison methods roughly doubled between the first and second half of the reviewed timeframe. Direct comparisons make up a quarter of all quantitative comparisons. Further, procedures in an oncological indication more often contain naive indirect comparisons than procedures in other therapeutic areas. Lastly, we find that in the investigated procedures, qualitative and direct comparisons are more often accepted as any type of indirect treatment

comparisons. This report shows that indirect comparisons play an important role in the assessment of orphan products and further work is needed to evaluate the appropriateness of different methodologies.

INTRODUCTION

In 2000, the Regulation (EC) No 141/2000 on orphan medicinal products became effective in the European Union. The legislation was introduced to incentivize development of medicinal products in populations affected by rare diseases, with the aim to ensure that treatments are also developed for patients with rare diseases. More than 20 years down the line there is clear evidence that the EU Orphan Regulation has made important contributions to overall development of new orphan medicines, 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 benefit to patients [1].

The regulation defines rarity as a condition not affecting more than 5 in 10,000 persons and further states 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]. Significant benefit can be defined either as a clinically relevant advantage or a major contribution to patient care (see Box 1). Significant benefit is assessed in comparison to all products approved for the therapeutic indication both at the time of initial orphan designation as well as at the time of marketing authorization as orphan medicinal product.

The criteria for demonstration of a significant benefit can be summarized as follows:
(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, e.g. 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 (e.g. 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 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, randomized 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 complicates 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 occasionally also 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 (for example different distributions of demographic characteristics). The available methods for indirect comparisons range from simple (unadjusted) methods like the side-by-side (SBS) comparison, over adjusted methods like the matching-adjusted indirect comparison (MAIC) [7] to more complex approaches taking into account whole networks of evidence of available treatments in a given indication (e.g. 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 furthermore determined by the reporting format of the different trials.
The Committee for Orphan Medicinal Products (COMP) is the central body responsible for evaluating applications for (maintenance of) orphan designation. It consists of experts representing all EU and EEA countries, as well as patient representatives, responsible for evaluating whether applicants fulfill the regulatory requirements for orphan designation, such as significant benefit. Anecdotal evidence and findings from a recent report suggested that indirect comparisons have been used more and more in recent years in support of the significant benefit at the time of marketing authorization for orphan medicinal products, and that more sophisticated methodologies like NMAs and MAICs were utilized [10].

To investigate this hypothesis, we conducted a systematic evaluation of the role of indirect comparisons in the context of demonstrating significant benefit for orphan medicines at the time of marketing authorization, 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 eleven years following the methodology described in the following section.

METHODS

study design and selection of 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 for all included procedures, direct or indirect comparisons against competitors on the market were necessary. To obtain an overview of the current 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 procedures pertaining to products with a marketing authorization 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 01/01/2012 and 31/12/2022 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 COMPs evaluation of each procedure. Therefore, there are 2 procedures which date back to 2011 in the data set, which are visible in all plots displaying time as a variable.

Procedures were included, irrespective of procedure type (initial marketing authorizations 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 (i.e. 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 (OMAR), 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 was 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 needed to be considered - on the procedure level and on 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 in the significant benefit section of the respective procedure, for each procedure. Importantly,

when a product was compared against the standard of care or best available therapy, 'best available therapy' or 'standard of care' were 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 identified in the significant benefit section, we recorded:

- information on the comparison method (e.g. NMA, MAIC etc.) and categorized the type comparison methods into the following categories:
    1. direct comparisons
    2. naive indirect comparisons
    3. adjusted indirect comparisons using aggregate data
    4. adjusted indirect comparisons using patient-level data
    5. qualitative comparisons

- main trial design and comparator trial design, and

- 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 procedures, whereas most analyses display frequencies relative to the absolute number of comparisons (i.e. all identified comparisons perform all procedures were merged).

definition of comparisons, trial designs and appraisal outcome

For this review, we have categorized all identified comparisons as follows. First, we distinguish between quantitative and qualitative comparisons. Qualitative comparisons describe those instances where a "satisfactory method" (in the following: comparator) was described as an adjunct treatment to the investigational product, or alternatively, where it was shown that there was no complete overlap in indications between comparator and investigational product. Quantitative comparisons, on the other hand, were categorized into direct and indirect comparisons. Direct comparisons included RCTs with an arm each for the investigational product and the comparator, as well as baseline comparisons, i.e., comparing the effect of the comparator product as measured at baseline against the effect of the intervention measured at the end of a trial. All indirect comparisons were further sub-categorized into three types. The methodologically most simple type is the SBS comparison, also called naïve comparison, where treatment effect data on the same outcome variable across two or more independent trials are extracted for both the investigational product and the comparator. The difference in summary statistics between the treatment of interest and the comparator (e.g. difference between objective response rates from the respective trials) is then evaluated without any adjustment or quantifying the comparison's uncertainty (e.g. by displaying a confidence interval). SBS comparisons are frequently also used to compare baseline characteristics. In contrast, all other indirect comparison methods, that used a formal hypothesis test and quantified the uncertainty of the estimated effect, were termed "inferential indirect comparisons" in analogy to the formal statistical inference they facilitate. These were further subdivided into a) methods using IPD for the comparator and b) methods using aggregate data for the comparator. Among the inferential indirect comparisons using IPD, we included NMAs, as well as other types of regression, either with or without the use of a matching method.

Among the inferential methods not using IPD for all arms, we included MAICs, NMAs using aggregate data, as well as simulated treatment comparisons and Bucher method comparisons. The outlined categorization was chosen to fit all identified comparisons, which is why qualitative comparisons were recorded, even though they were not the focus of this review.

The terms “main trial design” and “comparator trial design” used in this review describe the types of studies that were used as a basis for the comparisons, i.e., from which the data were extracted to perform the comparison between the investigational product and the approved product. The different trial designs were categorized as such for the purpose of this review:

- randomized controlled trial: all trials with multiple trial arms to which patients were randomly allocated;
- non-randomized trial: all trials with multiple trials arms, but non-randomized treatment allocation;
- single-arm trial (SAT): trials with a single (active) treatment arm;
- observational study: non-interventional studies that were not based solely on registry data;
- registry study: non-interventional studies specifically based on registry data;
- none: this label was used for all those qualitative comparisons which did not depend on trial data;
- multiple: this label was used for all aggregate data cited from multiple sources of literature;
- meta-analysis: the underlying design was categorized as such if the used data were pooled estimates extracted from meta analyses.

The COMP’s appraisal was categorized as follows: a comparison could either be accepted, rejected, or not considered. The latter means that the comparison was presented to the COMP as part of the applicant’s submitted documents, but no comment was made in the assessment report regarding the COMP evaluation of this comparison. Rejected comparisons were further categorized into the COMP’s specific evaluation of the clinical significance and the methodological soundness, respectively, if this could be discerned from the assessment report. Accordingly, a comparison could be categorized as ‘rejected’ based on either lacking clinical significance or methodological soundness alone, or because of a lack of both. Further, if this was not specified clearly in the assessment report, the rejected comparison was categorized as ‘rejected unclear’, in other words based on a global assessment. Lastly, we recorded cases as ‘unclear’ where multiple comparisons were presented between the investigational product and the comparator, but it could not be discerned which of the comparisons were considered relevant for the positive COMP decision.

statistical analysis

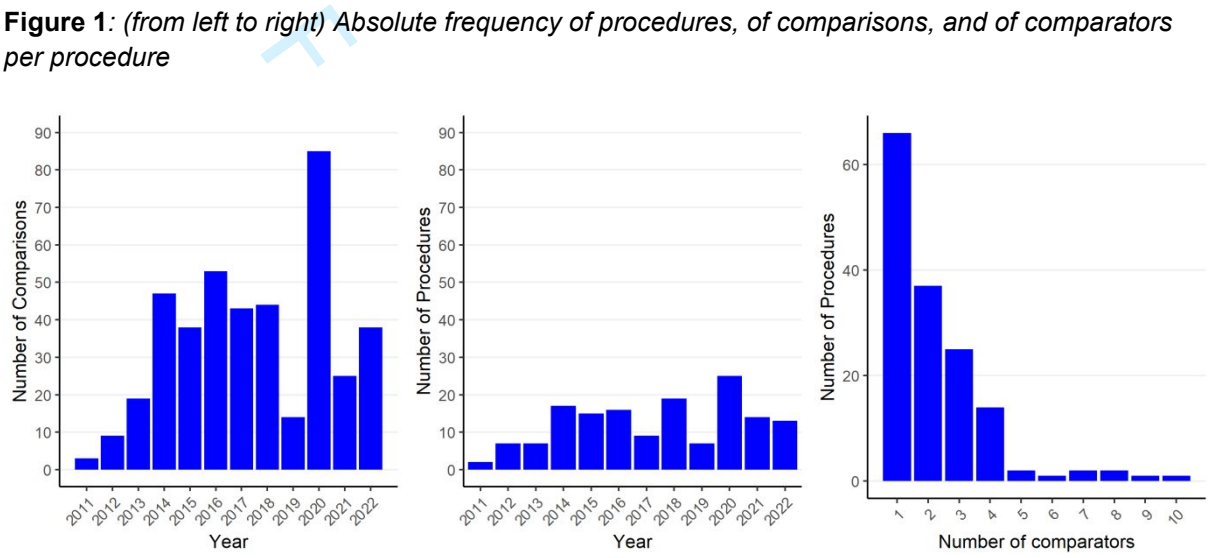
The data management and statistical analysis of all collected information was performed with R software [11], using the packages readxl, lubridate, tidyverse, ggplot2, scales, and reshape2. The main aim of the data analysis was to quantify the absolute and relative frequency of the use of different comparison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

Overall, 151 procedures were identified matching the inclusion and exclusion criteria. Within the specified timeframe, this was a subset of around 52% of all maintenance procedures (irrespective of outcome), and around 78% of all procedures which received a positive opinion, regardless of whether significant benefit had to be demonstrated or not. Across these 151 procedures, there were between 1 and 10 comparators per procedure (median = 3, interquartile range = 2 – 4; see Figure 1).

Figure 1: *(from left to right) Absolute frequency of procedures, of comparisons, and of comparators per procedure*



In roughly half of all cases, a list of questions was issued regarding the significant benefit. The final positive opinion was based on a clinically relevant advantage in the majority of procedures, but there were also several 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 table 1 in the appendix for an overview).

When looking at different disease areas targeted across procedures, using the system organ classes by the medical dictionary for regulatory activities categories [12], 60 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 28 procedures and 'Neoplasms benign, malignant and unspecified' with 18 procedures, where e.g. cystic fibrosis and ovarian cancer would be included respectively. Any other MedDRA categories were subject to 8 or less procedures (for an overview see Figure 1 in the appendix). More broadly, 68 of all 151 procedures were concerning an oncological indication, while 83 were non-oncology indications.

Overall, 418 comparisons were identified across all procedures (median = 2, interquartile range = 1 – 3, range = 1 – 14). 16 different types of comparison methods were identified, which were categorized into 5 broader groups of comparison types (see table 2).

Table 2

Category	Method	N	Short description
Quantitative, Direct comparisons			
	Head-to-head comparison	60	Direct comparison of two products as two parallel arms of one study, such as in a randomized controlled trial
	Baseline comparison	14	Comparing the outcome of one product measured at baseline of a study and the outcome of the other product at the end of the study
Quantitative, Indirect comparisons			
Side-by-side comparisons	Simple side-by-side comparison	113	Presentation of summary statistics for a variable (e.g. objective response rate for 'response') by treatment arms. The treatment arms are from separate studies, no statistical methods for cross-trial comparisons are applied (e.g. 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	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 [13].
	Bucher Method	7	Compares two or more products which have the same comparator (e.g. placebo) via indirect adjustment [14].
	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 [15].
	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 [16].
Inferential comparison with patient-level external data	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 a number of 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 (e.g. 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 [16].
Qualitative Comparison			
	Partial overlap in 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, e.g. 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

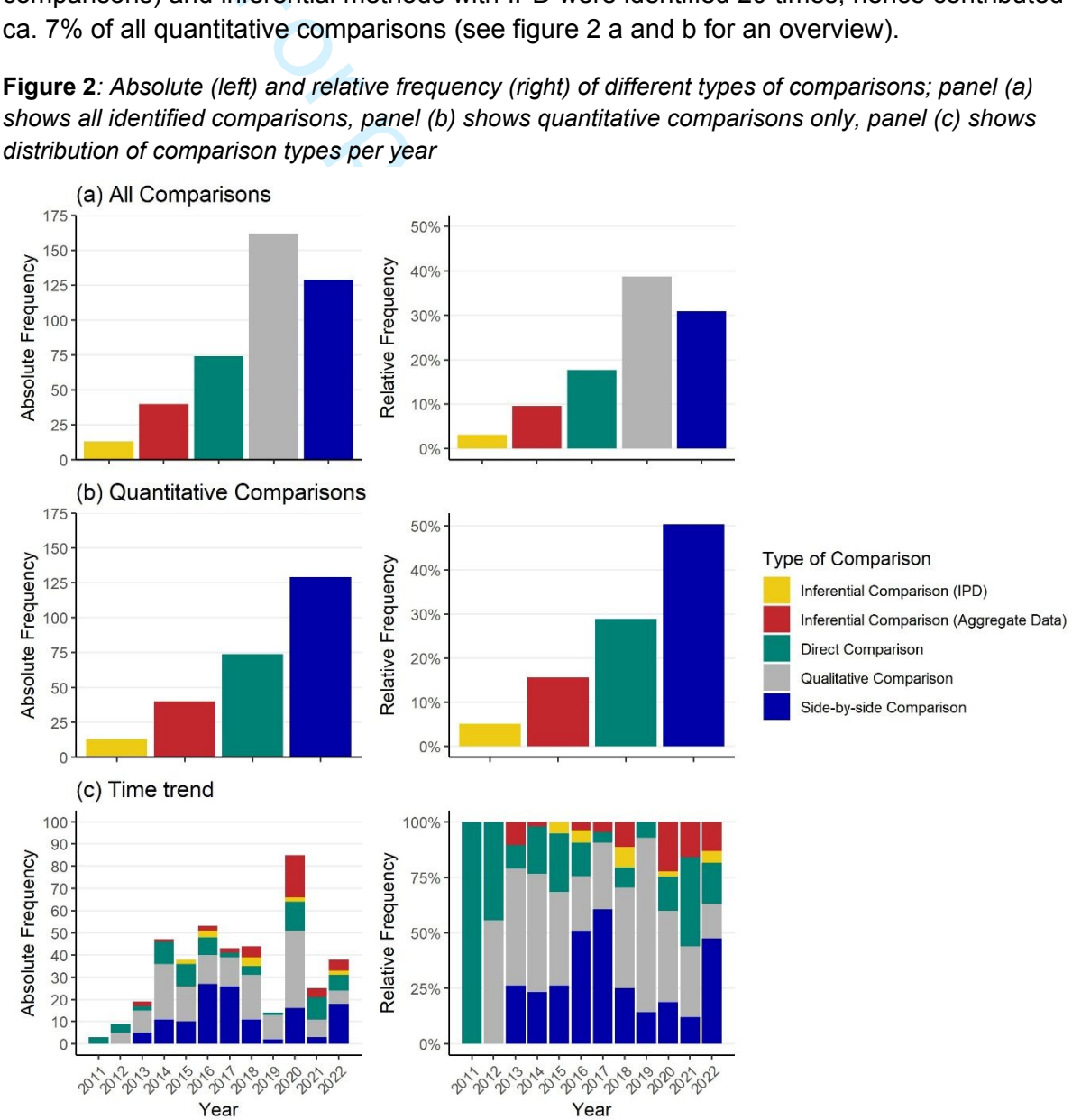
Table 2: Occurrence and short description of all comparison methods identified in the reviewed sample; the chosen categorization into 5 larger categories is reflected in all figures describing the identified comparisons and was chosen to reflect the most important methodological differences between the comparisons

When looking at the trial designs of the data sources underlying these comparisons, RCTs represented the majority with 284 of all main trials and 206 of all comparator trials. SATs were the next most frequent type of trial design, used as a source in 116 of all main trials and 28 of all comparator trials. For a full overview of trial designs see figure 2 in the appendix.

frequency of different comparison methods and development over time

We identified 74 direct comparisons, which constitutes 17% of all comparisons and 27% of all quantitative comparisons. 201 indirect comparisons were identified, corresponding to 73% of all quantitative comparisons. Of these indirect comparisons, SBS comparisons represented the majority (141, 51% of all quantitative comparisons), inferential methodologies utilizing aggregate data for the comparator arm were used 40 times (15% of all quantitative comparisons) and inferential methods with IPD were identified 20 times, hence contributed to ca. 7% of all quantitative comparisons (see figure 2 a and b for an overview).

Figure 2: *Absolute (left) and relative frequency (right) of different types of comparisons; panel (a) shows all identified comparisons, panel (b) shows quantitative comparisons only, panel (c) shows distribution of comparison types per year*



Comparing the first and second half of the investigated timeframe, between September 2011 and April 2017, 9% (17/190) of the identified comparisons were based on inferential methods (regardless of the use of IPD), whereas from May 2017 until December 2022, 17% (43/249) of the comparisons were based on inferential methods. Meanwhile, there was a slight decline in the proportion of direct comparisons from 19% (37/190) to 15%. When looking at SBS comparisons, 34% (64/190) were identified in the first half and 31% (77/249) in the second half of the reviewed timeframe (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 roughly doubled between the first and second half of the reviewed timeframe.

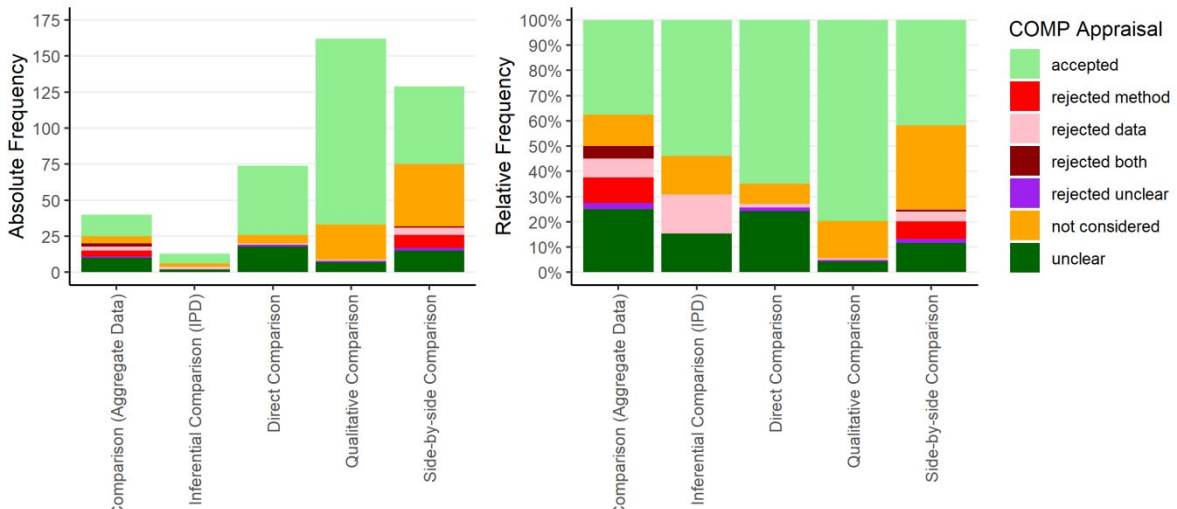
acceptance of different comparison methods by the COMP

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

62% (273/439) comparisons were accepted by the COMP, whereas 18% (81/439) were not considered and 8% (33/439) were not accepted. 12% (52/439) of comparisons to the approved products were deemed overall acceptable, leading to a positive opinion, but it could not be discerned which exact comparisons the COMP used as a basis for this decision. Among those not accepted, issues with the applied comparison methodology were the reason for the decision in 39% (13/33) of instances, while an insufficient effect difference was the reason for non-acceptance in 36% (12/33) of instances. In 9% (3/33) of the rejected comparisons, both methodology and absence of a sufficient benefit based on the magnitude of the relative effect were reasons for the rejection. 15% (5/33) of rejected comparisons were rejected with unclear grounds.

Investigating the acceptance of comparisons methods, the comparison method with the highest relative frequency of acceptance were 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).

Figure 3*: Absolute (left) and relative frequency (right) of the COMP's appraisal of comparisons*



To explore the appraisal of the different comparison methods by the COMP further, we also analyzed 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 which 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 procedures with a list of questions, (see Figure 4, top two panels).

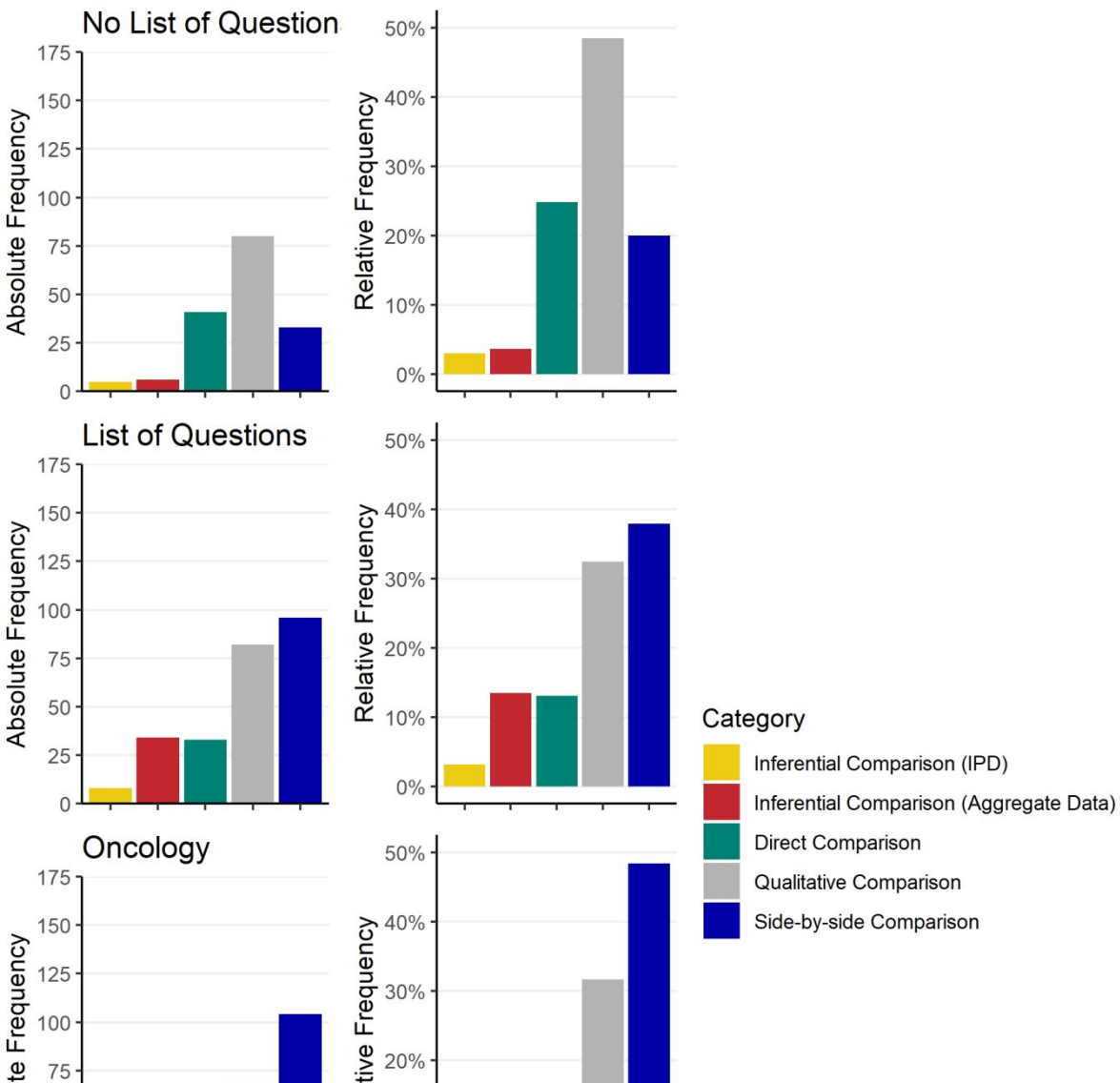
differences between therapeutic areas

To investigate potential differences between therapeutic areas regarding the choice of comparison methods, we distinguished all reviewed procedures into oncology and non-oncology procedures. While other categorizations would have been interesting to investigate as well, the distribution of therapeutic areas and the high proportion of oncology did not allow meaningful comparisons within the non-oncology indications.

Non-oncology procedures were supported by direct comparisons 2.5 times more often than oncology procedures, namely in 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 little over 10% in non-oncology procedures. The use of inferential indirect comparison methods, however, was higher in non-oncology procedures (for an overview, see Figure 4 below).

Further differences between oncology and non-oncology procedures can be seen regarding the trial design and appraisal of comparison method. SATs were the basis for comparisons far more often in oncology procedures than in non-oncology procedures (see Figure 3 in the appendix). Yet, RCTs were still the most used data source for main trials as well as comparator trials, in both non-oncology and oncology procedures. Looking at the COMP's appraisal, our data show that a lower proportion of comparisons was rejected in oncology procedures, particularly among all indirect comparisons (see Figure 4 in the appendix).

Figure 4: *Absolute (left) and relative frequency (right) of different types of comparisons across two stratifications, above procedures which in which a list of questions regarding the significant benefit was not or was issued, below showing all oncology procedures compared to all non-oncology procedures*



limitations

In this review, we have only included orphan maintenance procedures with a positive outcome. This choice was mainly driven by considerations on data accessibility. Most non-positive COMP opinions result in the applicant removing the orphan status voluntarily and progressing with a non-orphan marketing authorization. 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 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 for 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.

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 were performed, we have explored the types of approaches that have been used.

Overall, a high number of qualitative comparisons were used for demonstrating 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 driver 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 one to ten 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. While in oncology, SBS comparisons are the most-used method, for non-oncology products qualitative comparison followed by direct comparisons were most prominent. In fact, looking at the overall sample regardless of indication, 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 the overall acceptance of proposed comparisons was high, but differences between comparison approaches exist. While qualitative comparisons and direct comparisons were accepted in most cases, 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. In this context, it is important to highlight again that the acceptability of a comparison by COMP depends both on the comparison method and the data.

While the hypothesized 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 and more products have been approved for many rare diseases and the continued developments in network meta-analysis techniques, the importance of inferential statistical methods for indirect comparisons might further increase in the future.

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 authorization through the EMA CHMP, indirect comparisons can also play a role to demonstrate a major therapeutic advantage. After drug licensing, indirect comparisons play a crucial role for determining the relative effectiveness of authorized treatments as part of the health technology assessment. It would be interesting to explore similarities and differences between 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 authorization. While health technology assessment bodies regularly use and provide guidance on indirect comparison methods in order to compare the relative effectiveness of a new medicinal product [17, 18] 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 for helpful comments.

REFERENCES

[1] European Commission. Evaluation of the medicines for rare diseases and children legislation. 2020. 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. http://data.europa.eu/eli/reg/2000/141/oj

[3] Day S, Jonker AH, Lau LP, *et al.* Recommendations for the design of small population clinical trials. *Orphanet J. Rare Dis.* 2018 Dec;13:1-9. doi:10.1186/s13023-018-0931-2

[4] European Medicines Agency Committee for Human Medicinal Products. Guideline on clinical trials in small populations. 2006. 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 Dec;13:1-5. doi:10.1186/s13023-018-0926-z

[6] European Commission. Commission notice on the application of Articles 3, 5 and 7 of Regulation (EC) No 141/2000 on orphan medicinal products. 2016. 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 Sep 1;15(6):940-7. doi: https://doi.org/10.1016/j.jval.2012.05.004

[8] Dias S, Sutton AJ, Ades AE, Welton NJ. 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 Jul;33(5):607-17. doi: https://doi.org/10.1177/0272989X12458724

[9] Veroniki AA, Straus SE, Soobiah C, Elliott MJ, Tricco AC. A scoping review of indirect comparison methods and applications using individual patient data. *BMC Med. Res. Methodol*. 2016 Dec;16:1-4. doi: https://doi.org/10.1186/s12874-016-0146-y

[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 Aug 11;13:920336. doi: https://doi.org/10.3389/fphar.2022.920336

[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 Feb;20(2):109-17. doi: https://doi.org/10.2165/00002018-199920020-00002

[13] Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. *Pharmacoeconomics*. 2015 Jun;33(6):537-49. doi: https://doi.org/10.1007/s40273-015-0271-1

[14] Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J. Clin. Epidemiol*. 1997 Jun 1;50(6):683-91. doi: https://doi.org/10.1016/S0895-4356(97)00049-8

[15] Deeks JJ, Higgins JP, Altman DG, Cochrane Statistical Methods Group. Analysing data and undertaking meta-analyses. *Cochrane handbook for systematic reviews of interventions*. 2019 Sep 23:241-84. doi: https://doi.org/10.1002/9781119536604.ch10

[16] Chaimani A, Caldwell DM, Li T, Higgins JP, Salanti G. Undertaking network meta-analyses. *Cochrane handbook for systematic reviews of interventions*. 2019 Sep 23:285-320. doi: https://doi.org/10.1002/9781119536604.ch11

[17] Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Making*. 2018 Feb;38(2):200-11. doi: https://doi.org/10.1177/0272989X17725740

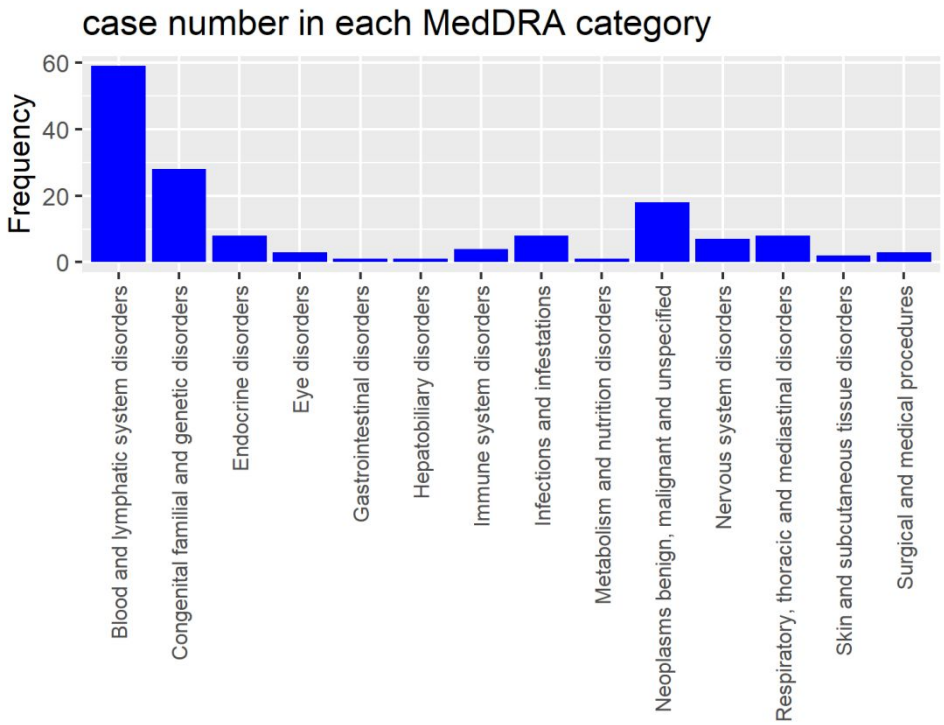
[18] EUnetHTA-21. Individual Practical Guideline Document D4.3.1: Direct and Indirect Comparisons. 2022. https://www.eunethta.eu/wp-content/uploads/2022/12/EUnetHTA-21-D4.3.1-Direct-and-indirect-comparisons-v1.0.pdf

Supplementary Materials

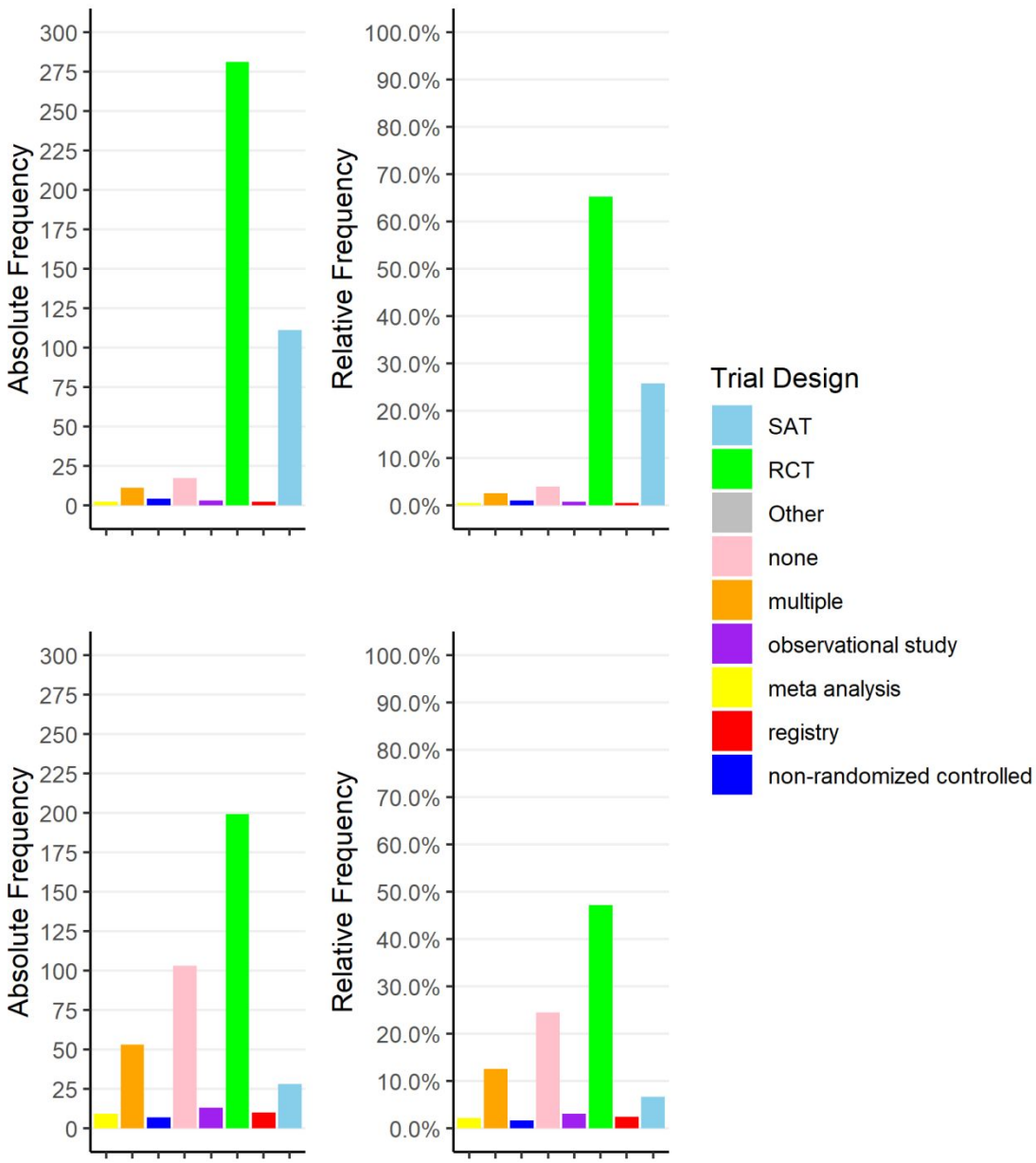
Category:			Number of occurrences:
All Maintenance procedures 2012-2022:			297
negative opinion:			8
withdrawn[1]			92
positive opinion:			197
positive opinion without significant benefit			46
positive opinion + significant benefit (**selected sample**):			151
	List of questions issued regarding significant benefit?	yes:	75
		no:	76
	Grounds for positive opinion:	clinically relevant advantage:	129
		major contribution to patient care:	15
		clinically relevant advantage + major contribution to patient care:	7
	Therapeutic area:	oncology:	68
		non-oncology:	83

Appendix Table 1: General characteristics of all identified procedures, as well as of the reviewed sample of procedures.

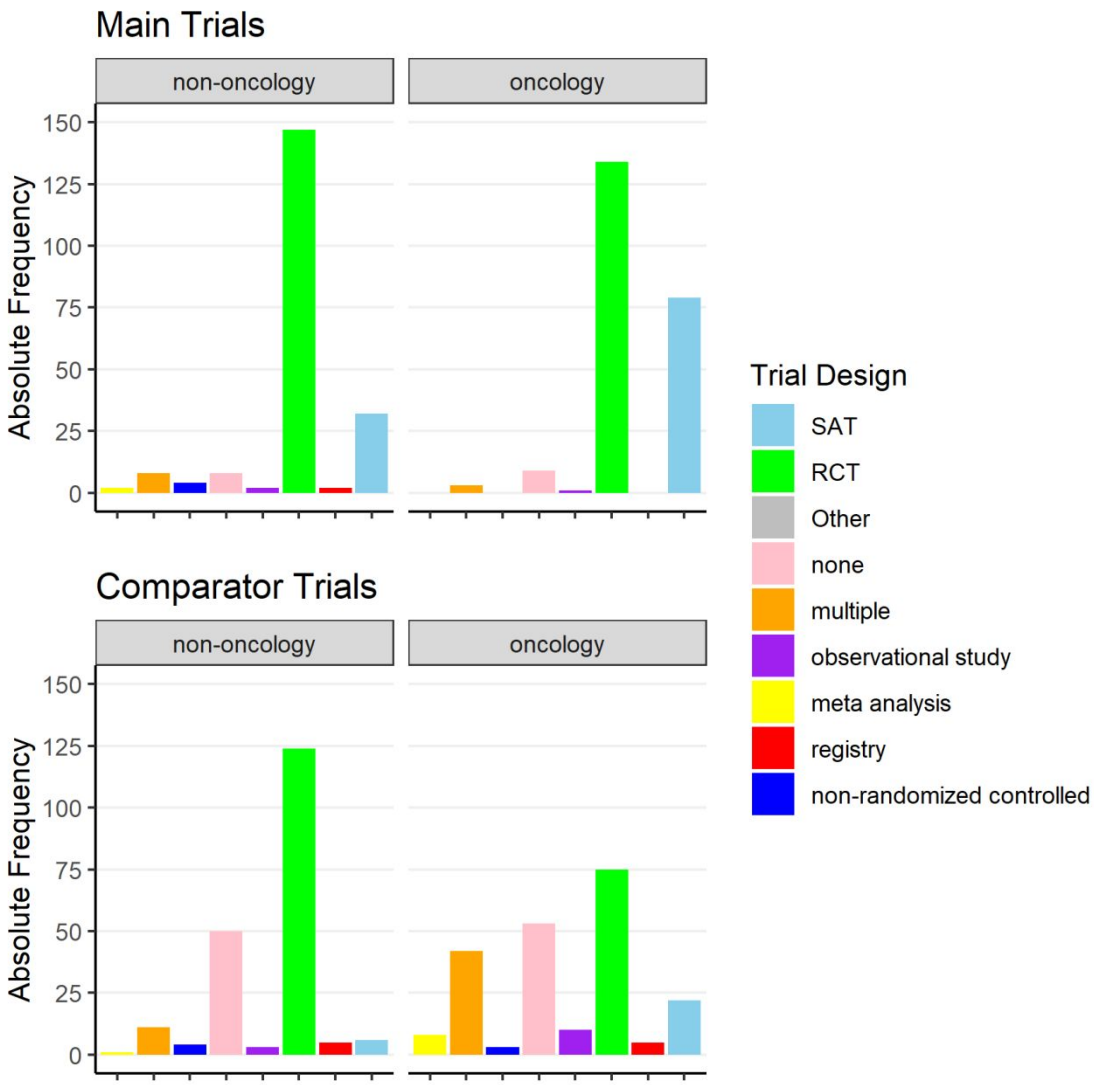
[1]There can be various reasons for withdrawals, such as if the designation holder anticipates a negative opinion from COMP, or if the designation holder received a negative opinion from CHMP, or if there are changes in the designation holder's regulatory strategy.



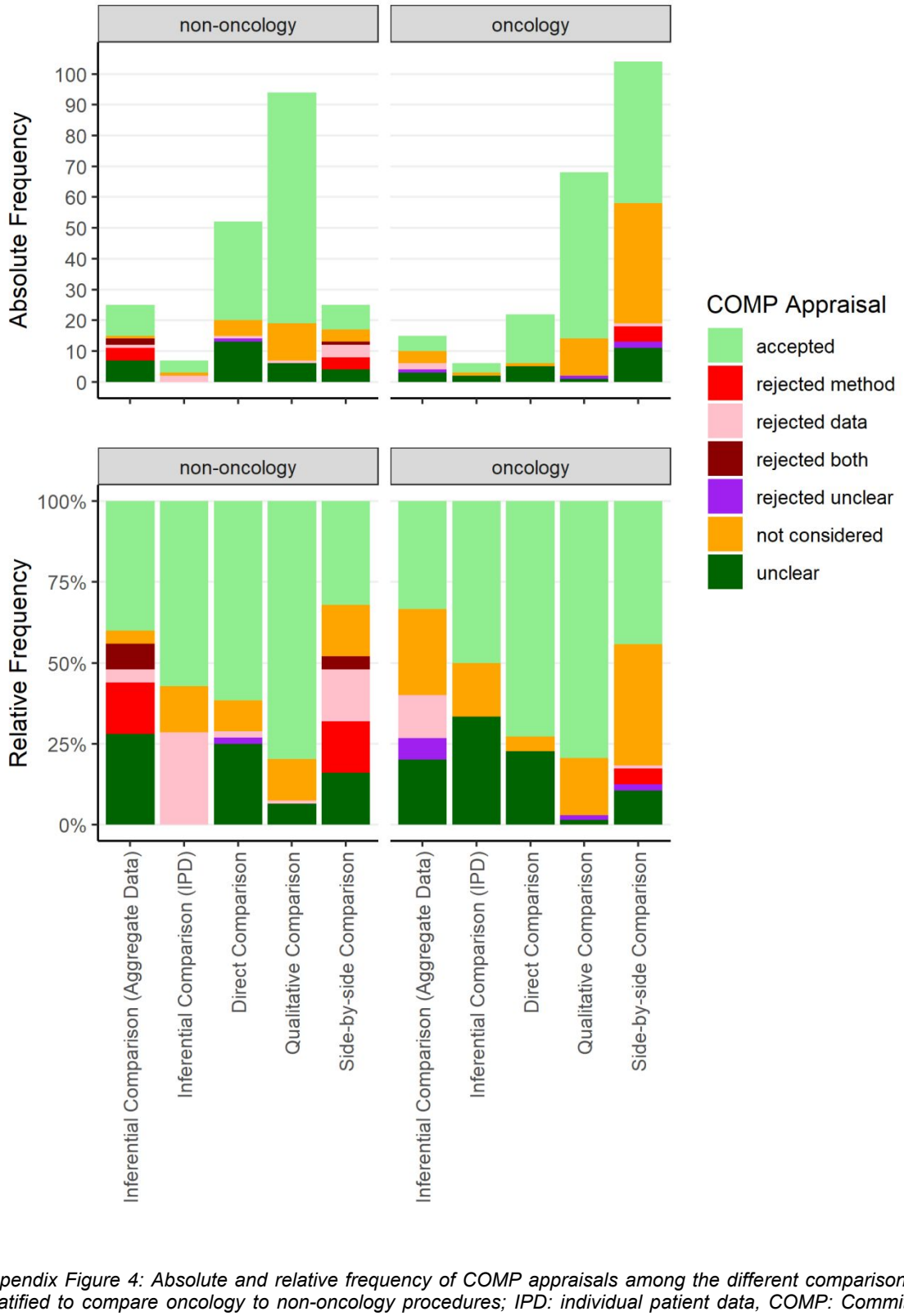
Appendix Figure 1: Number of analyzed procedures per disease area (categorized by System Organ Classes (SOCs) as featured in the Medical Dictionary for Regulatory Activities (MedDRA)



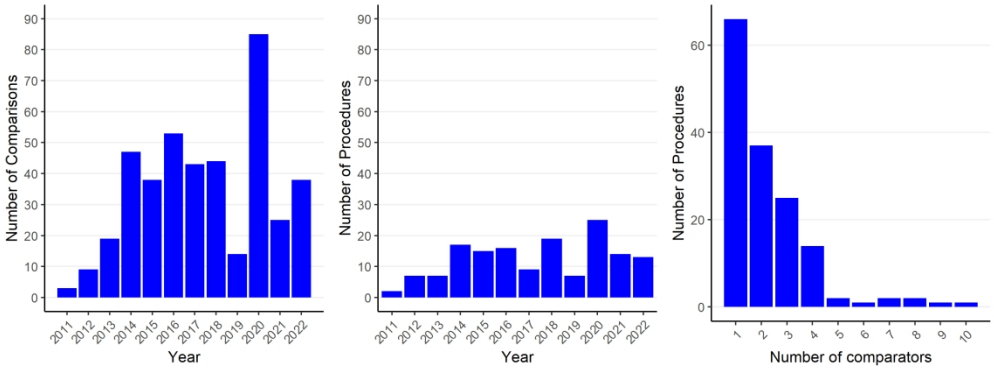
Appendix Figure 2: Absolute and relative frequency of identified trial design of the data underlying the effect of the investigational product and comparator product(s); RCT: randomized controlled trial, SAT: single-arm trial



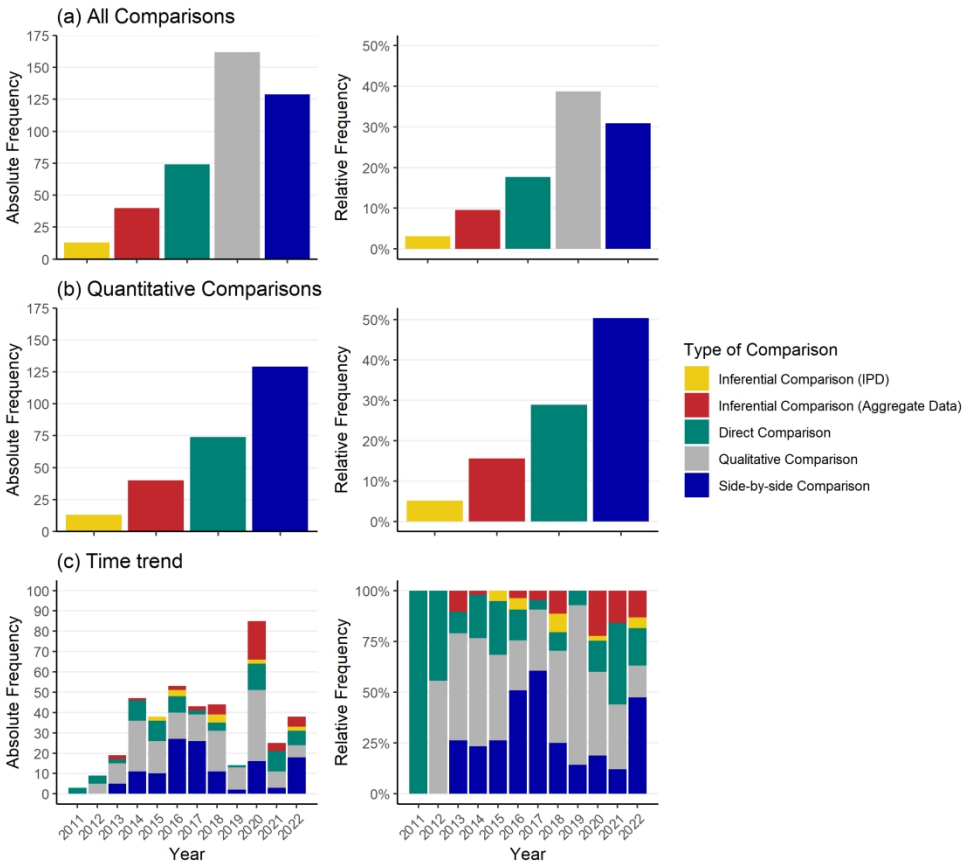
Appendix Figure 3: Absolute frequency of identified trial designs of oncology procedures compared to non-oncology procedures; RCT: randomized controlled trial, SAT: single-arm trial



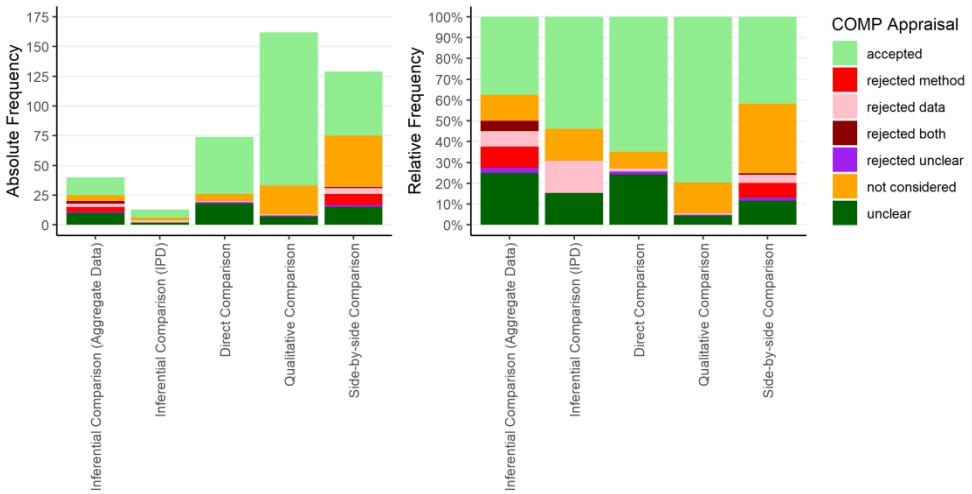
Appendix Figure 4: Absolute and relative frequency of COMP appraisals among the different comparison types, stratified to compare oncology to non-oncology procedures; IPD: individual patient data, COMP: Committee for Orphan Medicinal Products



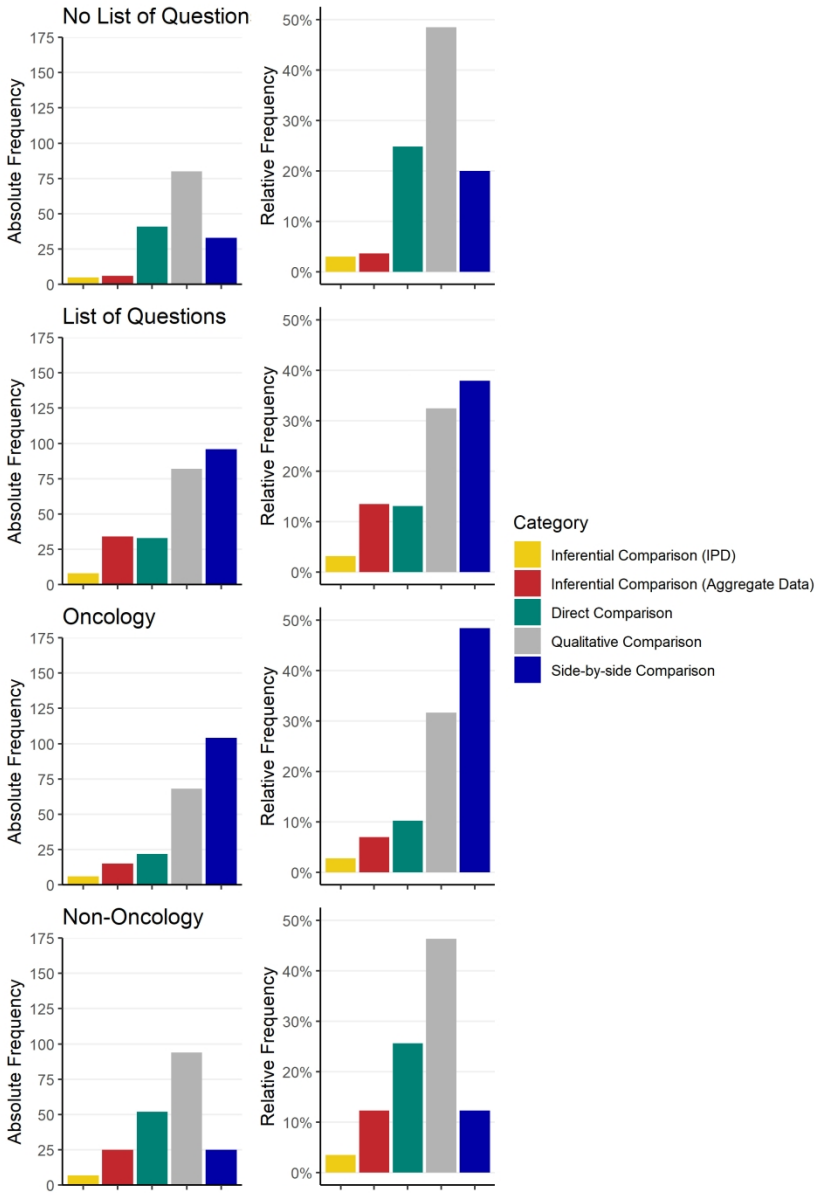
254x101mm (300 x 300 DPI)



228x203mm (300 x 300 DPI)



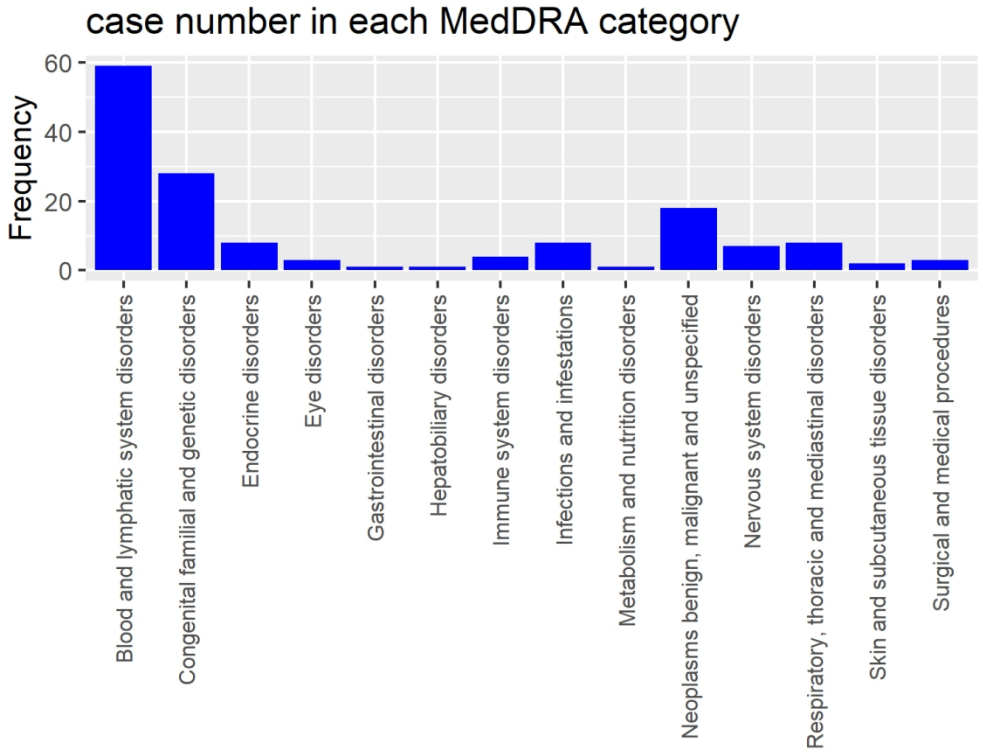
228x127mm (300 x 300 DPI)



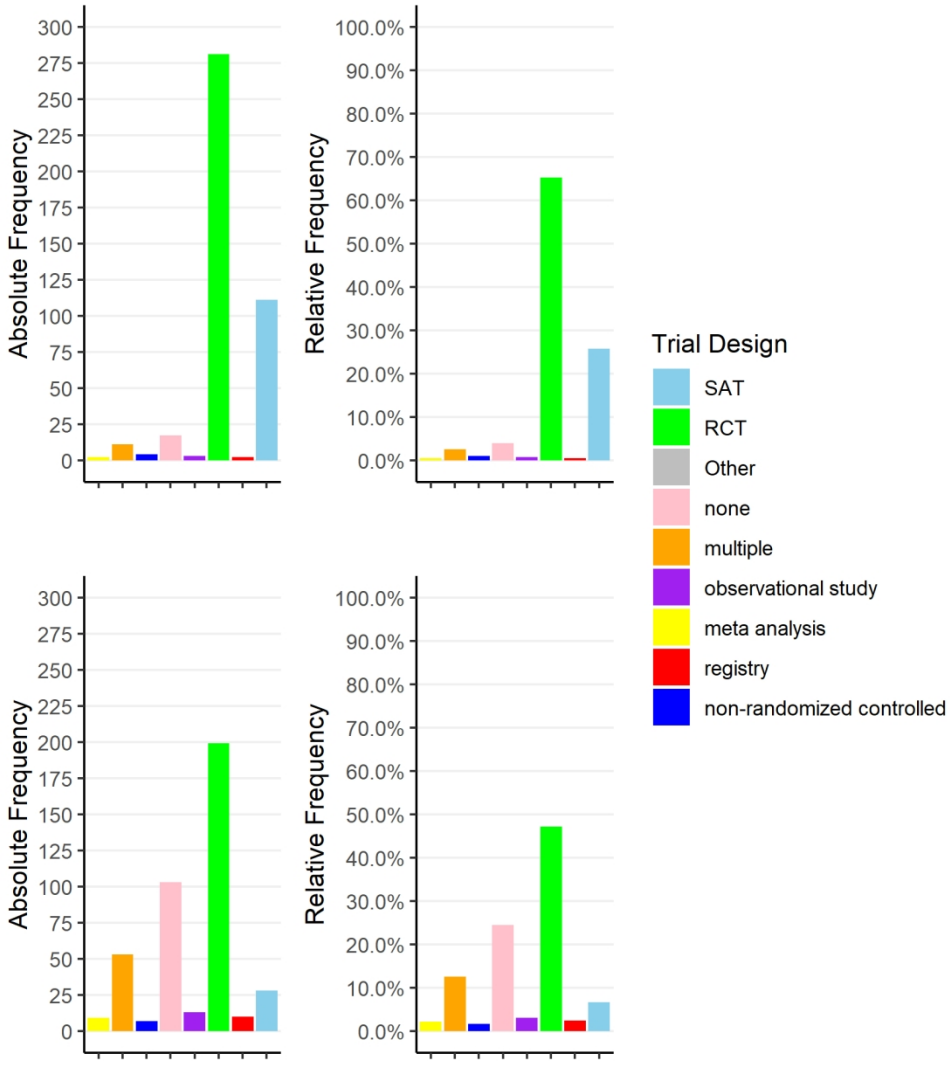
177x254mm (300 x 300 DPI)

Category:			Number of occurrences:
All Maintenance procedures 2012-2022:			297
negative opinion:			8
withdrawn[1]			92
positive opinion:			197
positive opinion without significant benefit			46
positive opinion + significant benefit (**selected sample**):			151
List of questions issued regarding significant benefit?		yes:	75
		no:	76
Grounds for positive opinion:		clinically relevant advantage:	129
		major contribution to patient care:	15
		clinically relevant advantage + major contribution to patient care:	7
Therapeutic area:		oncology:	68
		non-oncology:	83

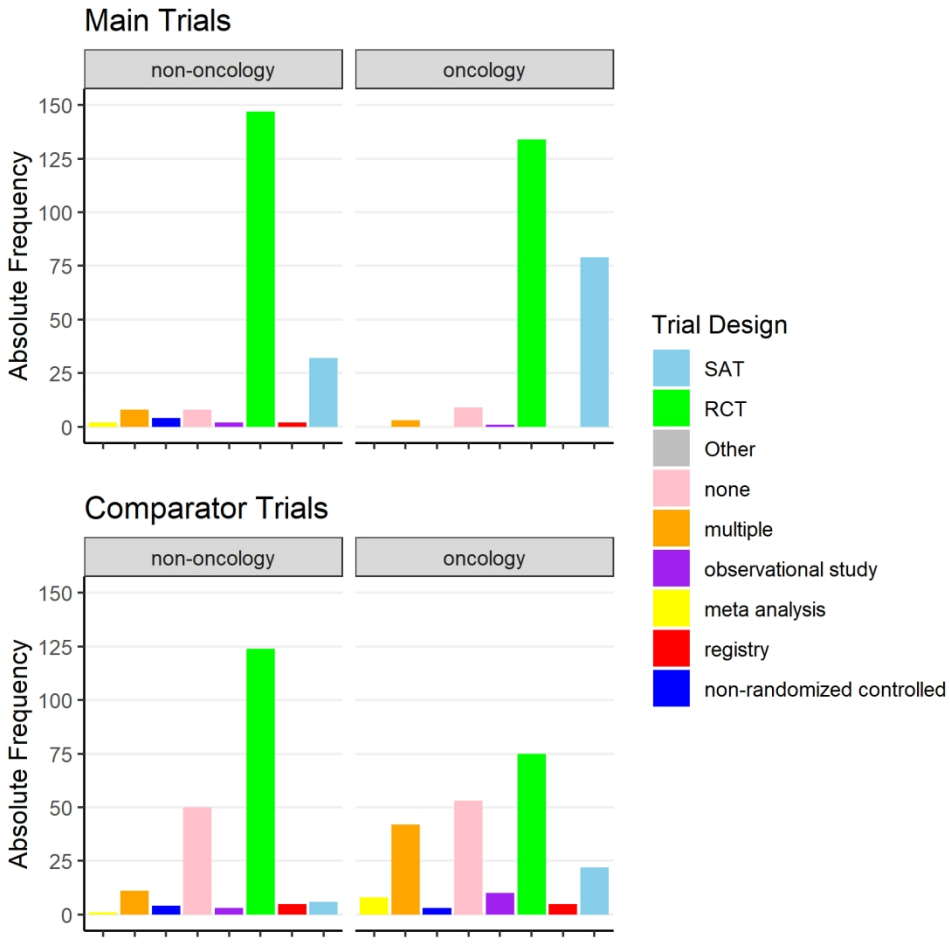
Appendix Table 1: General characteristics of all identified procedures, as well as of the reviewed sample of procedures.
[1]There can be various reasons for withdrawals, such as if the designation holder anticipates a negative opinion from COMP, or if the designation holder received a negative opinion from CHMP, or if there are changes in the designation holder's regulatory strategy.



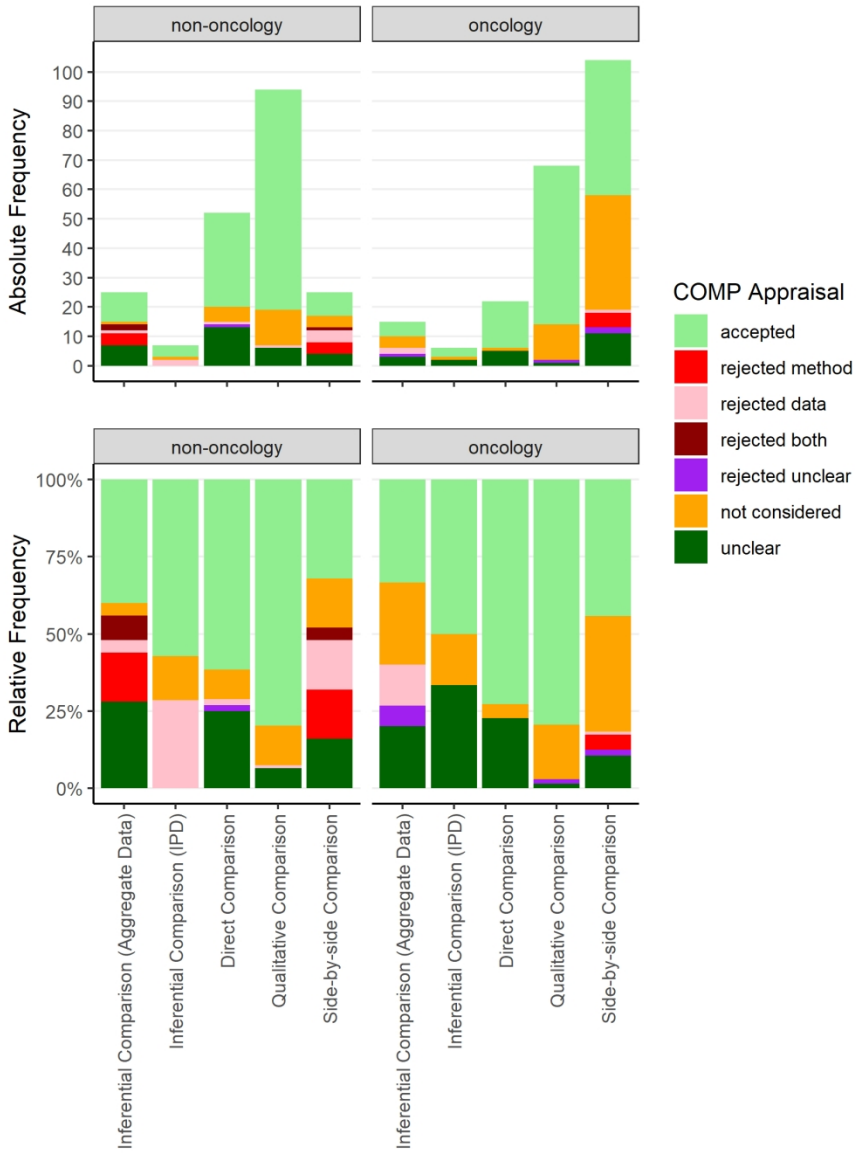
127x101mm (300 x 300 DPI)



152x177mm (300 x 300 DPI)



152x152mm (300 x 300 DPI)



152x203mm (300 x 300 DPI)

BMJ Open

Which clinical trial designs and statistical approaches have been used in assessments of orphan maintenance by EMA between 2012 and 2022 - a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-086171.R1
Article Type:	Original research
Date Submitted by the Author:	15-Oct-2024
Complete List of Authors:	Windfuhr, Fabian; European Medicines Agency, ; University Medical Centre Groningen Larsson, Kristina; European Medicines Agency Framke, Theodor; European Medicines Agency; Institute for Biostatistics Lasch, Florian ; European Medicines Agency; Institute for Biostatistics
Primary Subject Heading:	Health policy
Secondary Subject Heading:	Pharmacology and therapeutics, Public health, Research methods
Keywords:	Research Design, Clinical Trial, Drug Therapy, Legislation, Methods, Network Meta-Analysis

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Which clinical trial designs and statistical approaches have been used in assessments of orphan maintenance by EMA between 2012 and 2022 - a cross-sectional study.

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

[1] European Medicines Agency, Domenico Scarlattilaan 6, 1083 HS Amsterdam, The Netherlands
[2] University Medical Centre Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands
[3] Hannover Medical School, Institute for Biostatistics, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

Key words: *Orphan designation, significant benefit, indirect comparison, indirect treatment comparison, evidence synthesis.*

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.

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 EMA 'orphan maintenance' assessment documents between 2012 and 2022.

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. 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 less often used (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 roughly doubled between the first and second half of the reviewed timeframe.

conclusions:

Indirect comparisons play an important role in demonstrating a significant benefit in the assessment of orphan products and 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. The legislation was introduced to incentivize 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 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 benefit to patients [1].
In the European Union (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 (see Box 1). Significant benefit is assessed in comparison to all products approved for the therapeutic indication both at the time of initial orphan designation as well as at the time of marketing authorization of an orphan medicinal product. When a pharmaceutical company seeks an orphan designation, it is usually given at an early time point in 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 authorization on the other hand, it has to be assessed whether the orphan criteria are still met, i.e. "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 EEA member state, as well as three patient representatives, and additional topic experts. The COMP is responsible for evaluating whether applications fulfill the regulatory requirements for orphan designation, such as significant benefit.

The criteria for demonstration of a significant benefit can be summarized as follows:
(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, e.g. 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 (e.g. 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 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, randomized 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 complicates 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 occasionally also 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 (for example different distributions of demographic characteristics). The available methods for indirect comparisons range from simple (unadjusted) methods like the side-by-side (SBS) comparison, over adjusted methods like the matching-adjusted indirect comparison (MAIC) [7] to more complex approaches taking into account whole networks of evidence of available treatments in a given indication (e.g. 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 furthermore determined by the reporting 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 and more in recent years in support of the significant benefit at the time of marketing authorization for orphan medicinal products, and that more sophisticated methodologies like NMAs and MAICs were utilized [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 eleven years following the methodology described in the following 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 current 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 authorization 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 01/01/2012 and 31/12/2022 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 COMPs evaluation of each procedure. Therefore, there are 2 orphan maintenance procedures which 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 authorizations 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 (i.e. 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 (OMAR), 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 was 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 needed to be considered - on the procedure level and on 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' were 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 identified in the significant benefit section, we recorded:

- information on the comparison method and categorized the type comparison methods (see table 2 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 relative to the absolute number of comparisons. Details on the definition of the comparison, trial designs and appraisal outcome can be found in supplementary 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 received through this process was incorporated into the manuscript.

Statistical analysis

The data management and statistical analysis of all collected information was performed with R software [11], using the packages readxl, lubridate, tidyverse, ggplot2, scales, and reshape2. The main aim of the data analysis was to quantify the absolute and relative frequency of the use of different comparison methods, both 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 timeframe, 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 which received a positive opinion, regardless of whether significant benefit had to be demonstrated or not (see supplementary material 2 for more details). Across these 151 orphan maintenance procedures, there were between 1 and 10 comparators per procedure (median = 3, interquartile range = 2 – 4; see Figure 1).

Figure 1*: (from left to right) Absolute frequency of comparisons, orphan maintenance procedures, and of comparators per procedure*

INSERT FIGURE 1 HERE

In roughly half of all cases, a list of questions was issued regarding the significant benefit. The final 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 supplementary material 2 for an overview).

Using the system organ classes by the medical dictionary for regulatory activities categories [12] for categorizing 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 e.g. cystic fibrosis and ovarian cancer would be included respectively. Any other MedDRA categories were subject to 8 or less orphan maintenance procedures (for an overview see supplementary material 3). More broadly, 45% (68/151) of the orphan maintenance procedures were concerning an oncological indication.

Overall, 418 comparisons were identified across all of the 151 orphan maintenance procedures (median = 2, interquartile range = 1 – 3, range = 1 – 14). 16 different types of comparison methods were identified, which were categorized into 5 broader groups of comparison types (see table 2).

Table 2

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 randomized controlled trial
	Baseline comparison	14	Comparing the outcome of one product measured at 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 (e.g. objective response rate for 'response') by treatment arms. The treatment arms are from separate studies, no statistical methods for cross-trial comparisons are applied (e.g. 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 [13].
	Bucher Method	7	Compares two or more products which have the same comparator (e.g. placebo) via indirect adjustment [14].
	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 [15].
	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 [16].
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 a number of 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 (e.g. 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 [16].
Qualitative Comparison		162 (39%)	
	Partial overlap in 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, e.g. 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

Table 2: Occurrence and short description of all comparison methods identified in the reviewed sample; the chosen categorization into 5 larger categories is reflected in all figures describing the identified comparisons and was chosen to

reflect the most important methodological differences between the comparisons

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. SATs were the next most frequent type of trial design, 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 supplementary 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 2. Among the indirect comparisons naive side-by-side 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 individual patient data, are less often used (29%, 53/182).

Figure 2*: Absolute (left) and relative frequency (right) of different types of comparisons; panel (a) shows all identified comparisons, panel (b) shows quantitative comparisons only, panel (c) shows distribution of comparison types per year*

INSERT FIGURE 2 HERE

Comparing the first and second half of the investigated timeframe, between 2011 and 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) in the second half of the reviewed timeframe (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 roughly doubled between the first and second half of the reviewed timeframe.

Acceptance of different comparison methods by the COMP

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

The comparison method with the highest relative frequency of acceptance were 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 supplementary material 5).

Figure 3*: Absolute (left) and relative frequency (right) of the COMP's appraisal of comparisons*

INSERT FIGURE 3 HERE

To explore the appraisal of the different comparison methods by the COMP further, we also analyzed 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 which 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).

Differences between therapeutic areas

To investigate potential differences between therapeutic areas regarding the choice of comparison methods, we distinguished all reviewed orphan maintenance procedures into

oncology and non-oncology orphan maintenance procedures. While other categorizations would have been interesting to investigate as well, 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, namely in 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 little over 10% in 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 below).

Further differences between oncology and non-oncology orphan maintenance procedures can be seen regarding the trial design and appraisal of 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 vs 15%, 30/203 were SAT for the pivotal trial design in oncology and non-oncology, respectively, see supplementary material 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 supplementary material 8).

Figure 4*: Absolute (left) and relative frequency (right) of different types of comparisons across two stratifications, above orphan maintenance procedures which in which a list of questions regarding the significant benefit was not or was issued, below showing all oncology procedures compared to all non-oncology orphan maintenance procedures*

INSERT FIGURE 4 HERE

Limitations

In this review, we have only included orphan maintenance procedures with a positive outcome. This choice was mainly driven by considerations on data accessibility. Most non-positive COMP opinions result in the applicant removing the orphan status voluntarily and progressing with a non-orphan marketing authorization. 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 for 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).

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 for demonstrating 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 driver 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 one to ten 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 [17]. According to our data,, the proportion of rejected comparisons was lower in oncology compared to non-oncology indications. For context, prior research investigating the difference of 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 [18]. Meanwhile, it has also been reported that oncology products approved by EMA, often provide little or no added benefit [19], though no distinction between orphan or non-orphan product 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 benefit 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 hypothesized 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 and more products have been approved for many rare diseases and the continued developments in network meta-analysis 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 was available for the comparator that the new orphan medicine needed to compare against. 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 authorization through the EMA CHMP, indirect comparisons can also play a role to demonstrate a major therapeutic advantage. After drug licensing, indirect comparisons play a crucial role for determining the relative effectiveness of authorized treatments as part of the health technology assessment. It would be interesting to explore similarities and differences between 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 authorization. While health technology assessment bodies regularly use and provide guidance on indirect comparison methods in order to compare the relative effectiveness of a new medicinal product [20, 21] 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 & Author Contributions

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

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.

REFERENCES

[1] European Commission. Evaluation of the medicines for rare diseases and children legislation. 2020. 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. http://data.europa.eu/eli/reg/2000/141/oj

[3] Day S, Jonker AH, Lau LP, *et al.* Recommendations for the design of small population clinical trials. *Orphanet J. Rare Dis.* 2018 Dec;13:1-9. doi:10.1186/s13023-018-0931-2

[4] European Medicines Agency Committee for Human Medicinal Products. Guideline on clinical trials in small populations. 2006. 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 Dec;13:1-5. doi:10.1186/s13023-018-0926-z

[6] European Commission. Commission notice on the application of Articles 3, 5 and 7 of Regulation (EC) No 141/2000 on orphan medicinal products. 2016. 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 Sep 1;15(6):940-7. doi: https://doi.org/10.1016/j.jval.2012.05.004

[8] Dias S, Sutton AJ, Ades AE, Welton NJ. 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 Jul;33(5):607-17. doi: https://doi.org/10.1177/0272989X12458724

[9] Veroniki AA, Straus SE, Soobiah C, Elliott MJ, Tricco AC. A scoping review of indirect comparison methods and applications using individual patient data. *BMC Med. Res. Methodol*. 2016 Dec;16:1-4. doi: https://doi.org/10.1186/s12874-016-0146-y

[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 Aug 11;13:920336. doi: https://doi.org/10.3389/fphar.2022.920336

[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 Feb;20(2):109-17. doi: https://doi.org/10.2165/00002018-199920020-00002

[13] Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. *Pharmacoeconomics*. 2015 Jun;33(6):537-49. doi: https://doi.org/10.1007/s40273-015-0271-1

[14] Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J. Clin. Epidemiol*. 1997 Jun 1;50(6):683-91. doi: https://doi.org/10.1016/S0895-4356(97)00049-8

[15] Deeks JJ, Higgins JP, Altman DG, Cochrane Statistical Methods Group. Analysing data and undertaking meta-analyses. *Cochrane handbook for systematic reviews of interventions*. 2019 Sep 23:241-84. doi: https://doi.org/10.1002/9781119536604.ch10

[16] Chaimani A, Caldwell DM, Li T, Higgins JP, Salanti G. Undertaking network meta-analyses. *Cochrane handbook for systematic reviews of interventions*. 2019 Sep 23:285-320. doi: https://doi.org/10.1002/9781119536604.ch11

[17] Naci H, Davis C, Savović J, Higgins JP, Sterne JA, Gyawali B, Romo-Sandoval X, Handley N, Booth CM. 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 Sep 18;366. doi: https://doi.org/10.1136/bmj.l5221

[18] Rohr UP, Iovino M, Rudofsky L, Li Q, Juritz S, Gircys A, Wildner O, Bujar M, Bolte C, Dalla Torre di Sanguinetto S, Wolfer A. 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 and Translational Science*. 2023 Sep 16(9):1569-81. doi: https://doi.org/10.1111%2Fcts.13567

[19] Brinkhuis F, Goettsch WG, Mantel-Teeuwisse AK, Bloem LT. Added benefit and revenues of oncology drugs approved by the European Medicines Agency between 1995 and 2020: retrospective cohort study. *bmj*. 2024 Feb 28;384. doi: https://doi.org/10.1136/bmj-2023-077391

[20] Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Making*. 2018 Feb;38(2):200-11. doi: https://doi.org/10.1177/0272989X17725740

[21] EUnetHTA-21. Individual Practical Guideline Document D4.3.1: Direct and Indirect Comparisons. 2022. https://www.eunethta.eu/wp-content/uploads/2022/12/EUnetHTA-21-D4.3.1-Direct-and-indirect-comparisons-v1.0.pdf

Supplementary Materials

Supplementary Material 1

definition of comparisons, trial designs and appraisal outcome

For this review, we have categorized all identified comparisons as follows (see table 2 for detailed description). First, we distinguish between quantitative and qualitative comparisons. Qualitative comparisons describe those instances where a "satisfactory method" (in the following: comparator) was described as an adjunct treatment to the investigational product, or alternatively, where it was shown that there was no complete overlap in indications between comparator and investigational product. Quantitative comparisons, on the other hand, were categorized into direct and indirect comparisons. All indirect comparisons were further sub-categorized into three types. The methodologically most simple type is the SBS comparison, also called naïve comparison, where treatment effect data on the same outcome variable across two or more independent trials are extracted for both the investigational product and the comparator. The difference in summary statistics between the treatment of interest and the comparator (e.g. difference between objective response rates from the respective trials) is then evaluated without any adjustment or quantifying the comparison's uncertainty (e.g. by displaying a confidence interval). In contrast, all other indirect comparison methods, that used a formal hypothesis test and quantified the uncertainty of the estimated effect, were termed "inferential indirect comparisons" in analogy to the formal statistical inference they facilitate. The outlined categorization was chosen to fit all identified comparisons, which is why qualitative comparisons were recorded, even though they were not the focus of this review.

The terms "main trial design" and "comparator trial design" used in this review describe the types of studies that were used as a basis for the comparisons, i.e., from which the data were extracted to perform the comparison between the investigational product and the approved product. The different trial designs were categorized as such for the purpose of this review:

- randomized controlled trial: all trials with multiple trial arms to which patients were randomly allocated;
- non-randomized trial: all trials with multiple trials arms, but non-randomized treatment allocation;
- single-arm trial (SAT): trials with a single (active) treatment arm;
- observational study: non-interventional studies that were not based solely on registry data;
- registry study: non-interventional studies specifically based on registry data;
- none: this label was used for all those qualitative comparisons which did not depend on trial data;
- multiple: this label was used for all aggregate data cited from multiple sources of literature;
- meta-analysis: the underlying design was categorized as such if the used data were pooled estimates extracted from meta analyses.

The COMP's appraisal was categorized as follows: a comparison could either be accepted, rejected, or not considered. The latter means that the comparison was presented to the COMP as part of the applicant's submitted documents, but no comment was made in the assessment report regarding the COMP evaluation of this comparison. Rejected comparisons were further categorized into the COMP's specific evaluation of the clinical significance and the methodological soundness, respectively, if this could be discerned from the assessment report. Accordingly, a comparison could be categorized as 'rejected' based on either lacking clinical significance or methodological soundness alone, or because of a lack of both. Further, if this was not specified clearly in the assessment report, the rejected comparison was categorized as 'rejected unclear', in other words based on a global assessment. Lastly, we recorded cases as 'unclear' where multiple comparisons were presented between the investigational product and the comparator, but it could not be discerned which of the comparisons were considered relevant for the positive COMP decision.

Supplementary Material 2

Category:			Number of occurrences:
All Maintenance procedures 2012-2022:			297
negative opinion:			8
withdrawn[1]			92
positive opinion:			197
positive opinion without significant benefit			46
positive opinion + significant benefit (**selected sample**):			151
List of questions issued regarding significant benefit?		yes:	75
		no:	76
Grounds for positive opinion:		clinically relevant advantage:	129
		major contribution to patient care:	15
		clinically relevant advantage + major contribution to patient care:	7
Therapeutic area:		oncology:	68
		non-oncology:	83

Table: General characteristics of all identified procedures, as well as of the reviewed sample of procedures.
[1]There can be various reasons for withdrawals, such as if the designation holder anticipates a negative opinion from COMP, or if the designation holder received a negative opinion from CHMP, or if there are changes in the designation holder's regulatory strategy.

Supplementary Material 3

INSERT FIGURE 1 APPENDIX HERE

Figure: Number of analyzed procedures per disease area (categorized by System Organ Classes (SOCs) as featured in the Medical Dictionary for Regulatory Activities (MedDRA)

Supplementary Material 4

INSERT FIGURE 2 APPENDIX HERE

Figure: Absolute and relative frequency of identified trial design of the data underlying the effect of the investigational product and comparator product(s); RCT: randomized controlled trial, SAT: single-arm trial

Supplementary Material 5

Category:		N=418
Comparisons accepted by the COMP		253 (61%)
Comparisons not considered by the COMP		80 (19%)
Comparisons with unclear appraisal by the COMP		52 (12%)
Comparisons not accepted by the COMP		33 (8%)
	Not accepted due to methodological issues	13 (3%)
	Not accepted due to insufficient effect difference	12 (3%)
	Not accepted due to methodological issues and an insufficient effect difference	3 (1%)
	Not accepted on unclear grounds	5 (1%)

Table: absolute and relative frequencies of identified comparison appraisals by the COMP, as recorded in public assessment reports.

Supplementary Material 6

INSERT FIGURE 3 APPENDIX HERE

Figure: Absolute frequency of identified trial designs of oncology procedures compared to non-oncology procedures; RCT: randomized controlled trial, SAT: single-arm trial

Supplementary Material 7

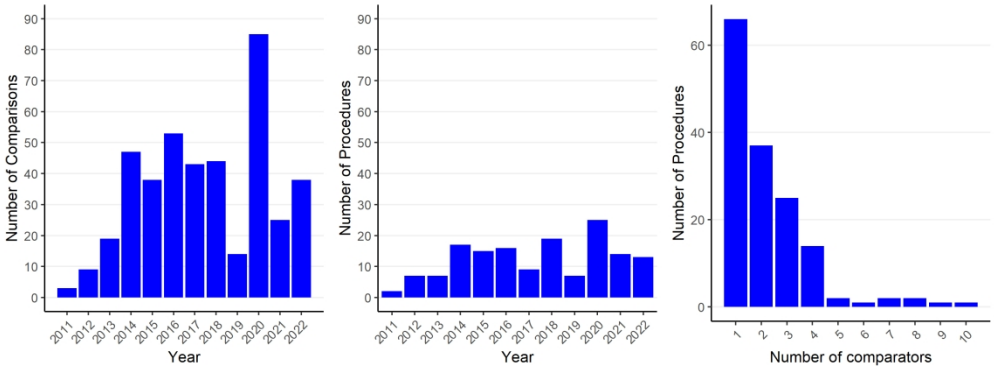
Main trial design	oncology (n=215)	non-oncology (n=203)
pivotal trial of new orphan medicinal product		
RCT	123 (57%)	145 (71%)
SAT	68 (32%)	30 (15%)
meta-analysis	0 (0%)	2 (1%)
non-randomised controlled	0 (0%)	4 (2%)
registry	0 (0%)	2 (1%)
observational study	1 (0.5%)	2 (1%)
multiple	14 (7%)	10 (5%)
none	9 (4%)	8 (4%)
pivotal trial of comparator medicinal product		
RCT	75 (35%)	123 (61%)
SAT	22 (10%)	5 (2%)
meta-analysis	8 (4%)	1 (0.5%)
non-randomised controlled	3 (1%)	4 (2%)
registry	2 (1%)	5 (2%)
observational study	7 (3%)	3 (1%)
multiple	45 (21%)	12 (6%)
none	53 (25%)	50 (25%)

table: design of pivotal and comparator trial for oncology and non-oncology orphan medicinal products

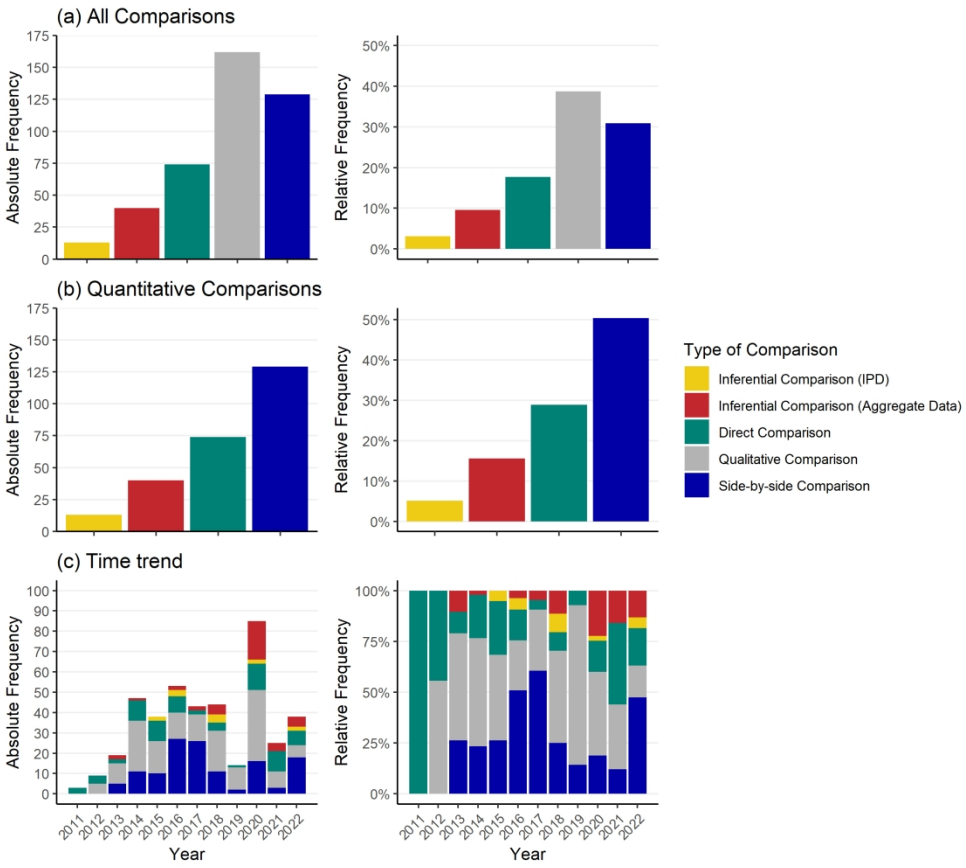
Supplementary Material 8

INSERT FIGURE 4 APPENDIX HERE

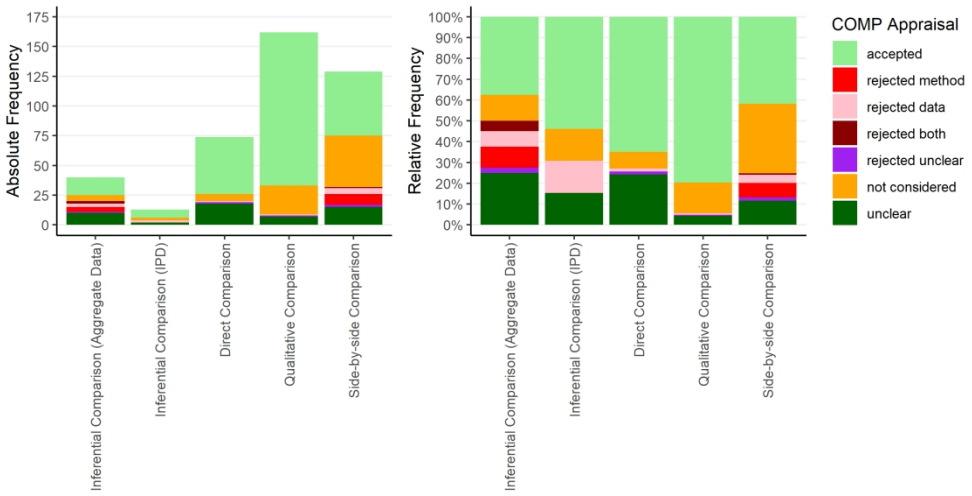
Figure: Absolute and relative frequency of COMP appraisals among the different comparison types, stratified to compare oncology to non-oncology procedures; IPD: individual patient data, COMP: Committee for Orphan Medicinal Products



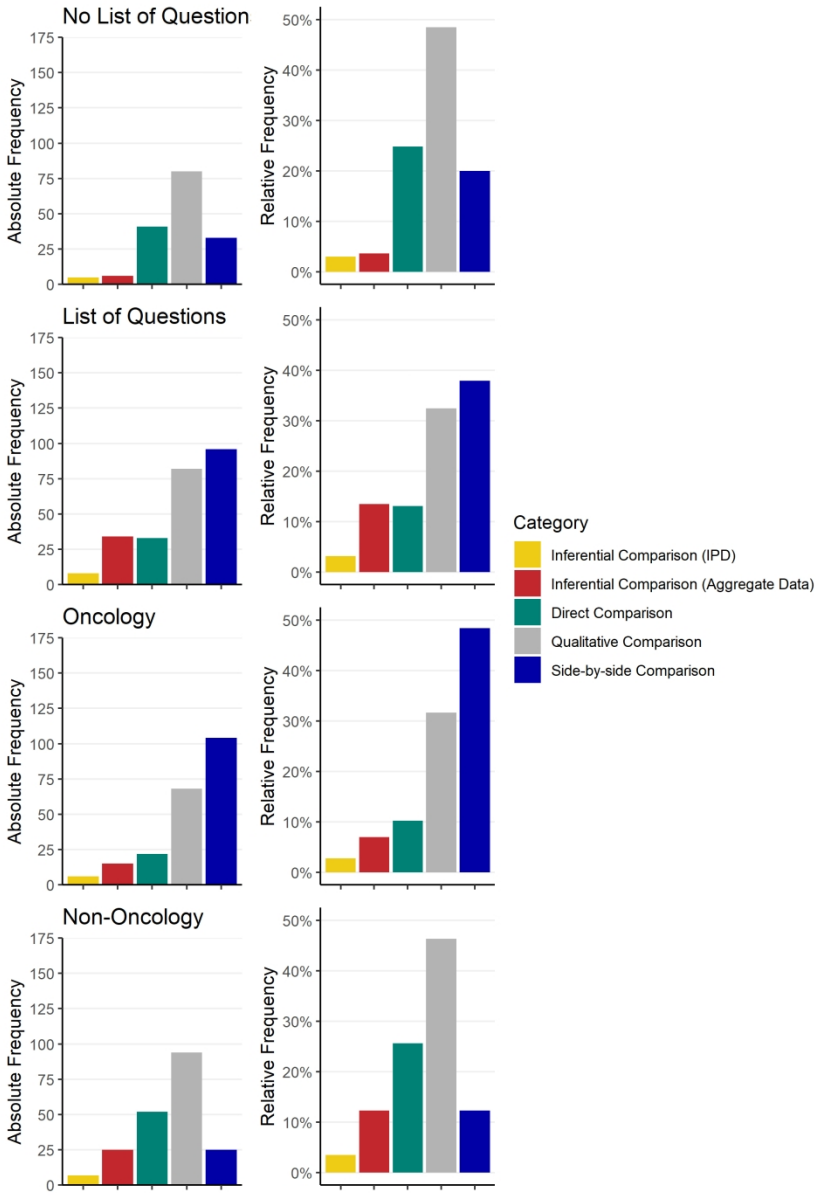
254x101mm (300 x 300 DPI)



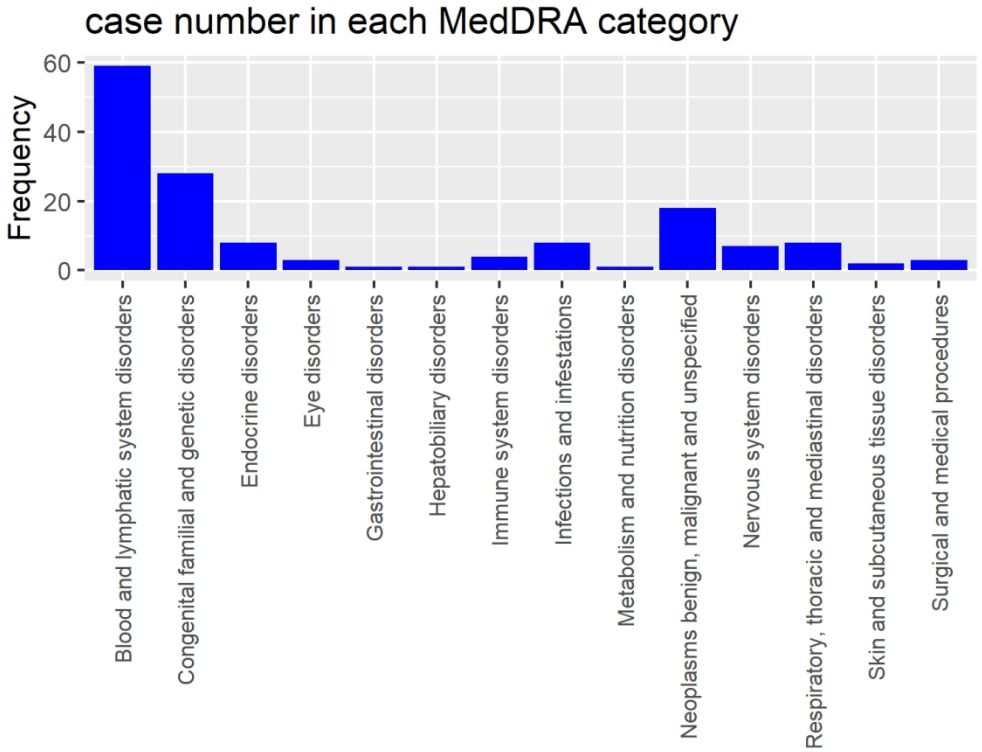
228x203mm (300 x 300 DPI)



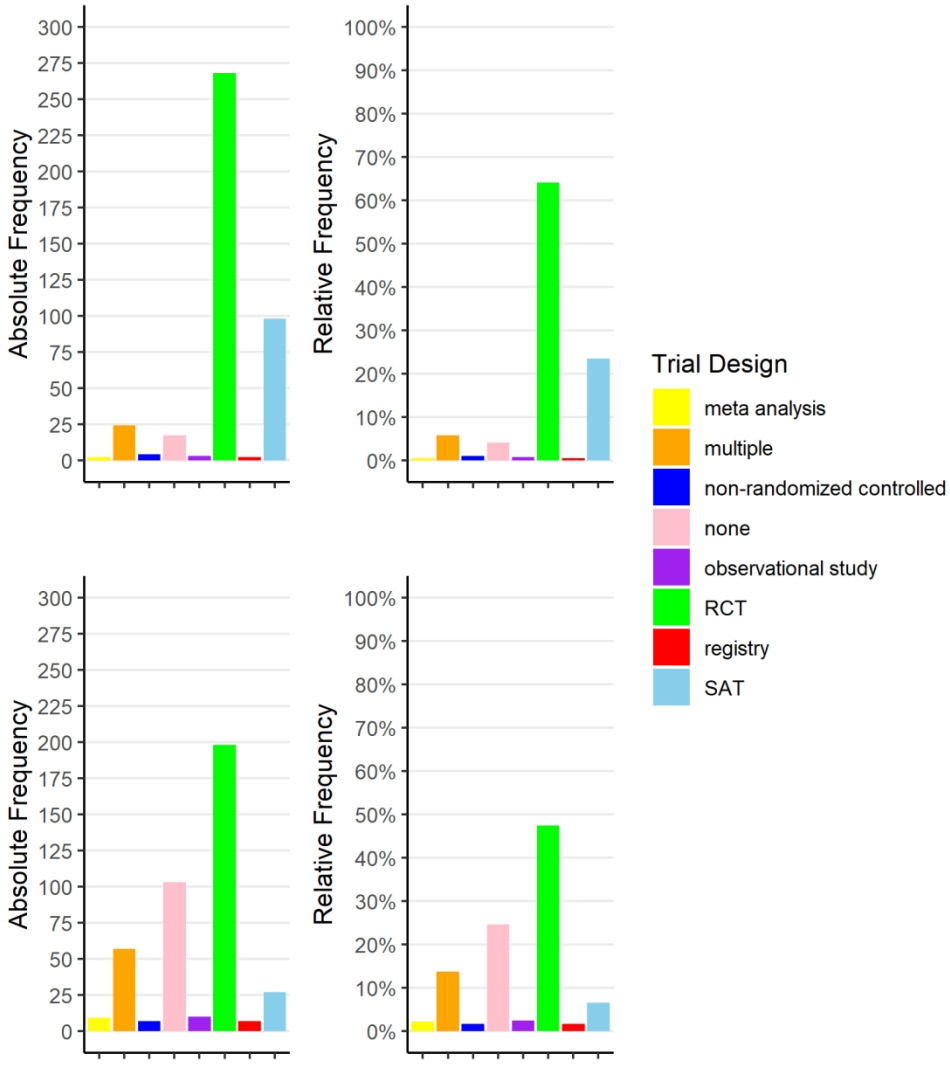
228x127mm (300 x 300 DPI)



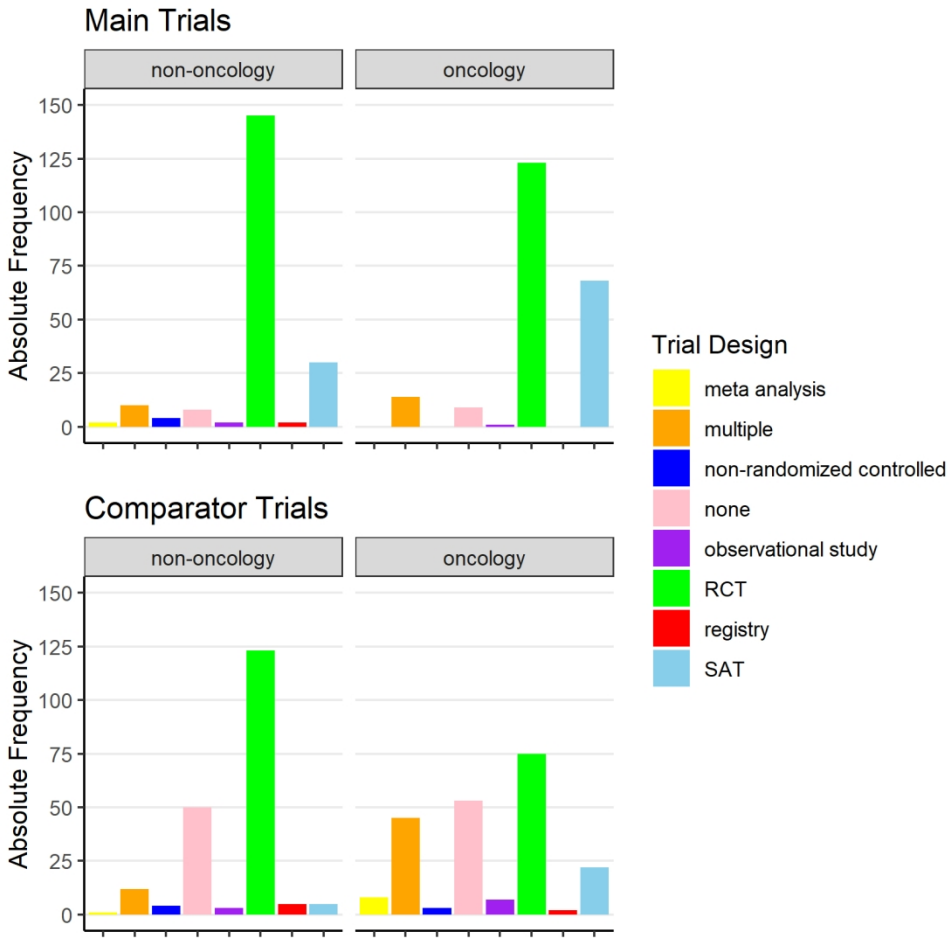
177x254mm (300 x 300 DPI)



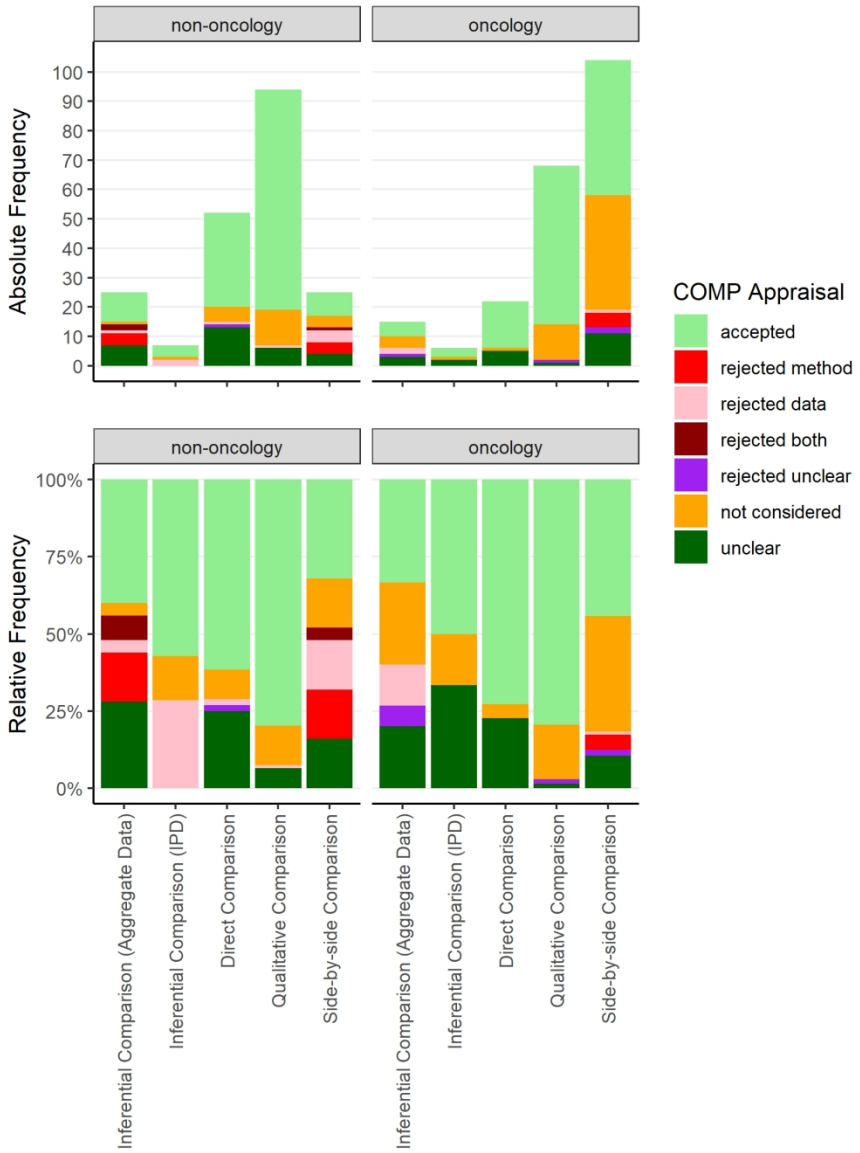
127x101mm (300 x 300 DPI)



152x177mm (300 x 300 DPI)



152x152mm (300 x 300 DPI)



152x203mm (300 x 300 DPI)