Variation in the Accuracy of Endpoint Selection in Clinical Studies for Rare Diseases in Respect of Increasing Knowledge on Disease-Severity Measurement

Ravi Jandhyala (ravi@medialis.co.uk)
King's College London

Research Article

Keywords: Neutral theory, clinical trial, rare disease, disease-severity measurement, accuracy, patient misclassification, trial recruitment

Posted Date: December 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1167543/v1

Abstract

Background: Previous research assessed the accuracy of disease-severity measurement in clinical studies as a mathematical relationship between the set of endpoints selected and the disease-severity scale (DSS), a surrogate for the theoretical Neutral list of indicators representing the disease phenotype. New DSSs are continually developed, so clinical studies' operationalisation of the Neutral list and resulting relative neutrality may vary over time. We assessed variation in the neutrality of clinical studies over time and the probability of false positive and false negative classifications at different disease prevalence rates.

Methods: We used search strings extracted from the Orphanet Register of Rare Diseases using a proprietary algorithm to conduct a systematic review of studies published until January 2021 per Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines. Overall, 483 studies and 12 rare diseases met inclusion criteria. We extracted all indicators from clinical studies and calculated neutrality and its components, sensitivity and specificity, as well as the probability of misclassifications at 20%, 50% and 80% disease prevalence rates at two time points, the times of publication of the first and last DSS. Surrogate Neutral lists were the first DSS and a composite of all later DSSs.

Results: Over time, the neutrality of clinical studies increased for six diseases and decreased for five diseases, driven by sensitivity for all but Friedreich ataxia. The neutrality of clinical studies in encephalitis decreased, but sensitivity remained constant at zero. At both timepoints, the likely false negative rate increased and the likely false positive rate decreased with increasing disease prevalence. The probability that the least neutral clinical study for most diseases would yield a false positive result was equal to one at all disease prevalence rates.

Conclusions: The potential for accurate clinical trial disease-severity measurement increases over time. Neutral theory showed that endpoint selection and DSSs may need improvement in Charcot Marie Tooth disease, Gaucher disease Type I, Huntington's disease, Sjogren's syndrome and Tourette syndrome. Using Neutral theory to benchmark disease-severity measurement in rare disease clinical trials may reduce the risk of misclassification, ensuring that recruitment and treatment effect assessment optimise medicine adoption and benefit patients.

Background

Ensuring the accuracy of clinical trial endpoints in rare diseases is important not only because trial results inform treatment decisions but also because disease progression for such conditions may remain poorly understood despite growing research attention.[1,2] Further, clinical trials for rare diseases are often small and have low power by default, so optimising measurement accuracy may ensure that endpoints translate to clinically meaningful results.[1] Clinical trials use distinct sets of indicators to observe disease severity, so the accuracy of disease severity measurement is affected by inter-trial variation,[3] which could be beneficial to address. This problem with inter-trial variation in accuracy has been quantified using

Neutral theory, which showed that when such sets of indicators are assessed against an aspirational list of all possible indicators of disease severity, one observer is likely to be more accurate than another.[4] This may result in variation in the number of patients classed as meeting a threshold of severity between clinical studies.[3,4]

Here, neutrality describes how accurately a set of clinical trial endpoints reflects the ideal construct for the disease being observed, which can be theoretically represented as an exhaustive list of disease severity indicators (the "Neutral list").[5] Neutrality can be expressed as the sum of a measure's sensitivity (number of relevant indicators excluded) and specificity (number of irrelevant indicators included), with a maximum score of two. Currently, the neutrality of disease severity measurement in rare disease clinical trial settings may range from zero to two, suggesting a need for improvement in both the sensitivity and specificity of disease severity measures used in clinical trials. This is especially salient in rare diseases, in which patients commonly show variation in disease phenotype[6] or may present atypically with severe disease in the absence of usual severity indicators.[7] While disease severity measures may correlate,[8] one study showed that as many as 20% of trial patients could have been excluded due to inter-trial variation in disease-severity measurement.[3]

Previous work on the neutrality of disease severity measurement has limited its scope to diseases with a single disease-severity scale (DSS) as a surrogate for the Neutral list for a disease. In this paper, we quantified the neutrality of disease-severity measurement in rare diseases with more than one DSS. This was necessary because severity measures are continually updated over time through scientific research,[9] so diseases have various DSS measures published at different times. In practice, if the neutrality of disease measurement in clinical trials varies over time, then the number of patients misclassified also varies at different time points. This has implications for the reliability of trial results and the decisions they inform. At a given time, the number of patients misclassified depends on the neutrality of the list of disease severity indicators used in clinical trials as a surrogate for the Neutral list.[5] Neutrality is also disease-specific, so while the Neutral list is a theoretical and unattainable ideal in practice, it can be used, in principle, to guide research streams to identify sets of indicators that are the best match for the Neutral list of a given disease at any given time. If the neutrality of disease measures changes over time, as the body of knowledge increases, then the best match for the Neutral list must be continually researched and updated as new evidence into disease-severity measurement becomes available.

To test this, we measured the neutrality of clinical studies at two time points, the time of publication of the first DSS and the time of publication of the most recent DSS, with the Neutral list surrogate at the second time point being a composite of all DSSs published up until that time. We assumed that the composite would reflect a more optimal match to the Neutral list, as it was based on a greater body of research, and new indicators of disease severity are explored continually as research advances.[11] Clinical studies were fixed compared to the first and composite DSS, so we expected the neutrality of clinical studies measured against them to decrease and the degree of patient misclassification to increase between the times of publication of the first and last DSS. In other words, we hypothesised that the

neutrality of clinical studies would vary as a function of the body of research conducted on the disease. Overall, this work focuses on the influence of time as a function of the wider body of knowledge done between two points and its impact on the neutrality of disease measurement and patient misclassification in clinical trials for rare diseases.

Methods

Aim

The aim was to assess variation in the neutrality of clinical studies over time and the probability of false positive and false negative classifications at different disease prevalence rates.

Design

A systematic review and statistical analysis were conducted.

Identification of clinical studies and rare diseases with more than one DSS

A previous study identified 26 rare diseases with at least one associated DSS by systematic review with search strings generated from all diseases listed on the Orphanet Register of Rare Diseases using a proprietary algorithm.[4] Of these, 15 were identified as having more than one DSS. A systematic review was conducted according to the preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidelines on the MEDLINE database, including all studies published until 21 January 2021. For each disease, the review used generic search strings as well as specific search strings generated from the Orphanet Register of Rare Diseases. All search strings have been provided in the appendix. To be included, studies had to be randomised controlled trials or observational clinical studies, conducted with human participants, have clearly recorded outcomes/endpoints, have full text available and be written in the English language. Animal and in-vitro studies were excluded. For a disease to be included in the study, there had to be at least two associated DSSs and more than five studies that used each of its associated DSSs to measure the severity of the disease in published literature. The studies also had to state clearly the disease studied. Two research analysts conducted the reviews, and the initial screening of titles and abstracts was conducted using Rayyan,[12] a web-based tool that automates systematic review processes.

Changes in the neutrality of clinical study disease measurement over time

The neutrality of clinical study disease measurement was assessed by first extracting all indicators reported in included clinical studies for each disease. Neutrality was defined as the sum of the sensitivity

and specificity of these sets of endpoints, where sensitivity was the proportion of relevant indicators included and specificity was the proportion of irrelevant indicators excluded as compared to the Neutral list.[5] In this study, we used the DSS as a surrogate for the Neutral list. As we expected that the Neutral list for each disease would vary over time as a function of the amount of research completed, we measured the neutrality of sets of clinical endpoints at two set points, the times of publication of the first and last DSS. Our surrogate Neutral list at the second time point was a composite of all indicators in all DSSs published at that time that met the inclusion criteria. We excluded duplicate indicators from the composites for each disease. We calculated the mean neutrality, sensitivity and specificity of measures used in clinical studies for each disease against the surrogate Neutral lists at each time point. Then, we calculated the neutrality, sensitivity and specificity of the most and least neutral clinical studies and computed the rates of false positive and false negative classifications these measures were likely to lead to at different disease prevalence rates according to previously described methods.[5] We observed neutrality at 20%, 50%, and 80% disease prevalence rates, representing clinical trial, observational study and clinical/outpatient settings, respectively. We treated sensitivity and specificity as statistically independent and included indicators from the composite DSS as part of the total information observed in the analysis for the first DSS (Table 1).

Table 1
Calculation of the sensitivity and specificity of clinical studies

	First DSS	Composite DSS
Sensitivity of clinical studies	Number of clinical study indicators present in the first DSS/(Total number of clinical study indicators + the total number of unique indicators in composite DSS)	Number of clinical study indicators present in the composite DSS/Total number of clinical study indicators
Specificity of clinical studies	(Total number of clinical study indicators – number of clinical study indicators absent from the first DSS)/(Total number of clinical study indicators + Number of unique indicators from the composite DSS)	(Total number of clinical study indicators – number of clinical study indicators absent from the composite DSS)/Total number of clinical study indicators

Results

Rare diseases, disease-severity scales and clinical studies included

Overall, 483 of the 2942 studies reviewed were included (Figure 1). Of the 26 diseases identified as having at least one associated DSS, 15 were identified as having more than one DSS. Gaucher disease Type 3 and Niemann Pick disease were identified in early screening as having more than one DSS but were eventually excluded due to insufficient published data on their disease-severity measures. Crohn's

disease also had more than one DSS, but research using the DSS did not distinguish between the disease measured, describing patients as having inflammatory bowel disease or a combination of both Crohn's and ulcerative colitis. Ulcerative colitis was excluded before review, as the previous study had identified it as having insufficient published data using its DSSs. The following 12 rare diseases were included (number of DSSs): acromegaly (3), amyotrophic lateral sclerosis (6), Charcot Marie Tooth disease (5), cystic fibrosis (5), encephalitis (4), Fabry disease (2), Friedreich ataxia (3), Gaucher disease Type 1 (2), Huntington's disease (2), juvenile rheumatoid arthritis (2), Sjogren's syndrome (2), and Tourette syndrome (3). An overview of the diseases, the composition and publication dates of their first and composite DSSs and the number of indicators in each has been provided in Table 2. Full details of indicators included in first and composite DSSs are available upon reasonable request. As shown in Table 1, the number of unique indicators available as the Neutral list for all diseases increased over time.

Table 2
Overview of diseases and first and composite disease-severity scales (year of publication)

Disease	First DSS [total number of indicators]	Composite DSS [total number of indicators, number of duplicates excluded]
Acromegaly	Clinical Activity Score of Acromegaly (1992) [17]	Clinical Activity Score of Acromegaly (1992) Acroscore (2015) ACRODAT® (2017) [26, 3]
Amyotrophic lateral sclerosis	Appel ALS Rating Scale (1987) [20]	Appel ALS Rating Scale (1987) Amyotrophic Lateral Sclerosis Severity Score (1989) Modified Norris Scale (1996) Amyotrophic Lateral Sclerosis Functional Rating Scale (1999) Amyotrophic Lateral Sclerosis Utility Index (2005) Japan ALS Severity Classification (2012) [71, 7]
Charcot Marie Tooth disease	Charcot-Marie-Tooth Neuropathy Score (2007) [9]	Charcot-Marie-Tooth Neuropathy Score (2007) CMT Neuropathy Score (2011) CMT Examination Score (2011) CMT Pediatric Scale (2012) Mobility-Disability Severity Index (2014) [26, 16]
Cystic fibrosis	The Shwachman-Kulczycki Score (1958) [20]	The Shwachman-Kulczycki Score (1958) Brasfield Score (1979) Cystic Fibrosis Clinical Score (1999) Matouk Clinical Score (2004) Chrispin–Norman Score (2005) [61, 10]

Disease	First DSS [total number of indicators]	Composite DSS [total number of indicators, number of duplicates excluded]
Encephalitis	The Status Epilepticus Severity Score (2008) [4]	The Status Epilepticus Severity Score (2008) END IT Score (2016) The Epidemiology-Based Mortality Score in Status Epilepticus (2015) The Clinical Assessment Scale in Autoimmune Encephalitis (2019) [43, 2]
Fabry disease	The Mainz Severity Score Index (2003) [25]	The Mainz Severity Score Index (2003) FD Severity Scoring System (2009) [37, 0]
Friedreich ataxia	The International Cooperative Ataxia Rating Scale (1997) [24]	The International Cooperative Ataxia Rating Scale (1997) The Scale for the Assessment and Rating of Ataxia (2004) Friedreich Ataxia Rating Scale (2010) [50, 6]
Gaucher disease type I	Gaucher Disease Severity Score Index – Type I (2008) [15]	Gaucher Disease Severity Score Index – Type I (2008) The Disease Severity Scoring System (2015) [25, 1]
Huntington's disease	The Unified Huntington's Disease Rating Scale Motor Score (2013) [7]	The Unified Huntington's Disease Rating Scale Motor Score (2013) Problem Behaviors Assessment for Huntington Disease (2015) [18, 0]
Juvenile rheumatoid arthritis	Clinical Disease Activity Index for RA (2005) [4]	Clinical Disease Activity Index for RA (2005) Juvenile Arthritis Disease Activity Score (2009) [8, 0]

Disease	First DSS [total number of indicators]	Composite DSS [total number of indicators, number of duplicates excluded]
Sjogren's syndrome	EULAR Sjögren's Syndrome Disease Activity Index (2010) [6]	EULAR Sjögren's Syndrome Disease Activity Index (2010) The Sjögren's International Collaborative Clinical Alliance Ocular Staining Score (2015) [11, 0]
Tourette syndrome	Yale Global Tic Severity Scale (1977) [15]	Yale Global Tic Severity Scale (1977) The Shapiro Tourette Syndrome Severity Scale Score (2004) Premonitory Urge for Tic Disorders Scale (2012) [28, 0]

Changes in the neutrality, sensitivity and specificity of clinical studies over time

Table 3 shows the mean change in neutrality, sensitivity and specificity of clinical studies over time when compared against the first and composite DSS as well as the number of clinical studies included for each disease. We found two main effects. First, for almost half the diseases (acromegaly, amyotrophic lateral sclerosis, cystic fibrosis, Fabry disease, and juvenile rheumatoid arthritis) the mean neutrality of clinical studies increased over time, and this appeared to be driven by an increase in sensitivity. The magnitude of increase in the mean neutrality of clinical studies for acromegaly (0.135), amyotrophic lateral sclerosis (0.094), Fabry disease (0.083), and juvenile rheumatoid arthritis (0.126) varied, with cystic fibrosis (0.022) showing the least variation. The sensitivity of clinical studies in most of these diseases increased threefold, but for juvenile rheumatoid arthritis, it was closer to a fivefold increase. Compared to this, the increase in the mean specificity of clinical studies for each of these diseases was negligible. For Friedreich ataxia, there was an increase in the neutrality of clinical studies, but the increase in sensitivity was much more modest than other increases in sensitivity and was of a similar magnitude to the increase in specificity.

Table 3
Average neutrality (sensitivity, specificity) of clinical studies measured against first and composite DSS

	Rare disease	Number of papers	Average neutrality	
			First DSS only	Composite DSS
1	Acromegaly	30	0.899 (0.029, 0.870)	1.034 (0.110, 0.923)
2	Amyotrophic lateral sclerosis	81	1.044 (0.109, 0.935)	1.138 (0.166, 0.972)
3	Charcot Marie tooth	12	1.371 (0.472, 0.898)	1.117 (0.215, 0.902)
4	Cystic fibrosis	156	0.991 (0.008, 0.983)	1.013 (0.026, 0.987)
5	Encephalitis	12	0.956 (0.000, 0.956)	0.905 (0.000, 0.905)
6	Fabry disease	15	0.935 (0.039, 0.896)	1.018 (0.105, 0.913)
7	Friedreich ataxia	18	1.102 (0.299, 0.803)	1.248 (0.344, 0.904)
8	Gaucher type I	15	0.969 (0.142, 0.851)	0.943 (0.112, 0.831)
9	Huntington's disease	32	1.456 (0.500, 0.956)	1.162 (0.205, 0.957)
10	Juvenile rheumatoid arthritis	39	0.983 (0.032, 0.951)	1.109 (0.147, 0.962)
11	Sjogren's syndrome	37	1.323 (0.401, 0.922)	1.146 (0.224, 0.922)
12	Tourette's syndrome	36	1.784 (0.837, 0.946)	1.395 (0.448, 0.946)

Second, for almost half the diseases (Charcot Marie Tooth disease, Gaucher disease Type I, Huntington's disease, Sjogren's syndrome, and Tourette syndrome), the mean neutrality of clinical studies decreased over time, and this appeared to be driven by a decrease in sensitivity. The magnitude of decrease in the mean neutrality for Charcot Marie Tooth disease (-0.254), Huntington's disease (-0.294), Sjogren's syndrome (-0.177), and Tourette syndrome (-0.389) varied, with Gaucher disease Type I (-0.026) showing the least variation. The sensitivity of clinical studies for most of these studies halved, but for Gaucher disease Type I, the decrease was much more modest. As was true for the other group of diseases discussed above, compared to the decrease in sensitivity, the decrease in specificity was negligible. There was a small decrease in the mean neutrality of clinical studies for encephalitis from the first to the composite DSS, but sensitivity remained constant at zero, while specificity decreased by a modest degree like that shown by the other diseases.

Changes in potential false positive and false negative classifications arising from changes in the neutrality, sensitivity and specificity of clinical studies over time

Figures 2 and 3 show the potential false positive and false negative classifications arising from changes in the neutrality, sensitivity and specificity of clinical studies over time when assessed against the first and composite DSS, respectively. First, overall, for both first and composite DSSs, the probability of a false negative result increased with increasing disease prevalence, while the probability of a false positive decreased. Second, for both first and composite DSSs, the probability that the least neutral clinical study for most diseases would yield a false positive result was equal to one at all disease prevalence rates. However, for encephalitis, this was true for both the most and least neutral study. There were no instances of the probability of a false negative equalling one for either the first or composite DSS.

Discussion

There is a pressing need to improve the accuracy of clinical trial endpoints in rare diseases.[1] Previous research has addressed this by assessing the neutrality of clinical endpoints in rare diseases using the DSS as a surrogate for the Neutral list.[4] However, some diseases had more than one DSS, meaning that the neutrality of any study measured against them may vary over time as a function of the body of knowledge on a disease. We expected that the neutrality of the fixed sample of clinical studies would decrease over time, as the body of knowledge (operationalised here as the number of indicators) increased. The number of indicators in the surrogate Neutral list for all diseases increased over time, but we found that the neutrality of clinical studies for one subgroup of diseases increased over time, while it decreased for another subgroup. This suggested that the neutrality of clinical studies changed in different ways with respect to the body of knowledge.

In the first subgroup of diseases, the mean neutrality of clinical studies was higher when measured against the composite DSS than the first DSS, and this appeared to be driven by an increase in sensitivity. That is, as the number of indicators in the surrogate for the Neutral list increased over time as a function of growth in the body of knowledge, clinical studies included a greater proportion of its indicators. This was not merely a function of the higher number of indicators increasing the probability of a match between the composite DSS and clinical studies, because we did not find this pattern across all diseases. Rather, this suggested a convergence of knowledge whereby DSS indicators generated through scientific research over time also showed up in the group of clinical studies assessed. We assumed that the sample of clinical studies would be static in terms of growth in the body of knowledge; however, it contained research spanning many years, so the body of knowledge that informed the composite may also have informed a proportion of the clinical studies in the sample. In the study of scientific epistemology (for example, Peirce's convergence of truth and the mathematics upon which it is based), a convergence of knowledge on a construct observed alongside an increase in sample size can be taken as a sign of the validity of the knowledge generated.[13,14]

The convergence of knowledge between DSSs and clinical studies in the first subgroup suggested that the disease phenotype operationalised by the indicators shared between them tended towards a more accurate representation of the theoretical Neutral list over time, producing more accurate measures of disease severity. In the second subgroup of diseases, we observed a decrease in neutrality, which we expected under the incorrect assumption that clinical studies would be unaffected by the increasing body of knowledge. Given our findings in the first subgroup, this can be better interpreted as a divergence of knowledge. If the clinical and DSS studies were methodologically sound, then this divergence may be fertile ground for hypothesis building and further knowledge generation.[15] Further research may examine qualitative differences between indicators in each disease, as our findings suggested that DSSs in the first subgroup were more likely to contain indicators that were specific, measurable and objective and that were pathophysiological as well as behavioural and psychological. As we measured clinical studies as a homogenous group and did not separate them out into two timepoints, our findings cannot suggest that the changes in neutrality found represented a specific relationship between neutrality and time within the diseases studied. However, our methods were sufficient to demonstrate that changes in how the Neutral list is operationalised over time affect the accuracy of clinical trial disease measurement and that this must be accounted for during the selection of endpoints.

The mean sensitivity of clinical studies for encephalitis remained constant at zero, suggesting that clinical studies included no indicators relevant to disease severity as defined by the DSS at either time point, suggesting a divergence of knowledge regarding the operationalisation of disease phenotype between clinical studies and DSSs. The Clinical Assessment Scale for Autoimmune Encephalitis is the only disease-specific DSS developed and validated for use in patients with encephalitis.[16,17] Other DSSs for encephalitis included here were designed and validated to measure status epilepticus, a single intracranial complication that only covers part of the disease phenotype.[18–20] This could account for the lack of overlap of indicators between clinical studies and DSSs at both timepoints. Additionally, heterogeneity of disease phenotype between and within rare disease subtypes, as is found to a great degree in encephalitis,[21–23] may affect the neutrality and standardisation of disease-severity measurement in clinical trials.[9]

The inaccurate measurement of disease severity in clinical trials may result in patient misclassification.[3,9,10] We measured the impact of neutrality via its components, sensitivity and specificity, on the probability of detecting false negative and false positive results at different disease prevalence rates. In a clinical trial setting (20% prevalence rate), in many diseases, the probability of a false positive was equal to one (the classification of a patient as ‘severe’ when they are ‘not severe’). If these disease-severity measures were used as inclusion criteria for trials, our findings suggested a high probability of including patients outside of the target population. Additionally, in many diseases, specificity was equal to zero, meaning that all indicators observed in clinical studies were irrelevant to disease severity. The detection of a treatment effect in these cases could result in the licensing of a medicine with little clinical significance to patients. If no treatment effect was detected, then trials may be abandoned, and effective medicines may be rejected at the regulatory stage, meaning that potentially life-

changing medications may fail to reach patients, which is a recurrent problem in rare disease clinical trials and may be attributed to lack of neutrality in endpoint selection.[24,25] Further, for these diseases, outcomes of relevance to disease severity may be underrepresented in the body of research, so patients may not benefit from ongoing evidence generation regarding the problems they deal with in their day-to-day lives. We observed a similar pattern of data at all prevalence rates that became more pronounced as prevalence increased. This was in line with previous findings and gave confidence in our results.[4]

Limitations

First, we assumed that the DSS was a surrogate for the Neutral list, as it was the most accurate representation of the disease phenotype available. However, the Neutral list is an empirically unattainable theoretical concept.[5] This is likely to have resulted in an over-estimation of the neutrality of clinical studies in this study than if a 'true' measure of neutrality was made. Second, Neutral theory assumes that indicators are independent of each other; however, associations may exist between indicators to varying degrees. Finally, we did not control for the effect of time of publication of clinical studies, which may be reasonably expected to affect the number of indicators they shared with the surrogate Neutral list to some degree (clinical studies published before composite DSSs may be less likely to contain their indicators, although this is not guaranteed, as DSSs are generated based on existing bodies of knowledge shared by those who conduct trials). The variation in the year of publication of DSSs between diseases was not suggestive of a confound in respect of the effects noted in this study, and most DSSs were published between 5 and 10 years before the analysis.

Conclusions

Our results suggested that the potential for accuracy in measuring disease severity increases as a function of the body of knowledge on a disease. The neutrality of almost half the rare diseases in this study increased as the body of knowledge increased, while the neutrality of almost half decreased, suggesting that sustained research efforts in some diseases resulted in the development of more accurate measures of disease severity implemented in DSSs and clinical studies. The application of Neutral theory could enhance the accuracy of endpoint selection in clinical trials and verify the accuracy and relevance of treatment effects as well as ensuring that the risk of misclassification during trial recruitment and the assessment of treatment effects is kept as low as possible. Further research may be beneficial to develop more accurate disease-severity measurements in Charcot Marie Tooth disease, Gaucher disease Type I, Huntington's disease, Sjogren's syndrome and Tourette syndrome.

Abbreviations

DSS Disease-severity score

Declarations

Ethical Approval

This study was exempt from ethical approval as the study authors collected and synthesized data from previous publications, in which informed consent had already been obtained by the primary investigators, per Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines. No human participants were recruited for this study.

Consent for Publication

Not Applicable.

Availability of data and materials

All data generated or analysed during the current study are included in this published article and are available from the corresponding author on reasonable request.

Competing interests

The author is a visiting senior lecturer at the Centre for Pharmaceutical Medicine Research at King's College London and is responsible for research into real-world evidence approaches. He is also the founder and CEO of Medialis Ltd, a medical affairs consultancy and contract research organisation involved in the design and delivery of real-world evidence in the pharmaceutical industry.

Funding

The author received no funding for this work.

Author contributions

RJ conducted the study and developed and approved the manuscript. The author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted, and any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Acknowledgements

The author would like to acknowledge the contribution of Medialis research analysts (Mohammed Kabiri, Shivshankar Sundar and Sandra Harry) and medical writer (Lauri Naylor) to the systematic review and

manuscript writing. Former Medialis employee, Solomon Christopher, led the statistical analysis.

Transparency Declaration

Medialis Ltd supports the development of accurate disease-severity tools for use in clinical studies as part of its Corporate Social Responsibility Programme.

References

1. Cox, GF. The art and science of choosing efficacy endpoints for rare disease clinical trials. Am J Med Genet Part A. 2018;176:759–772. https://doi.org/10.1002/ajmg.a.38629
2. Selvadurai, L.P., Georgiou-Karistianis, N., Shishegar, R. et al. Longitudinal structural brain changes in Friedreich ataxia depend on disease severity: the IMAGE-FRDA study. J Neurol. 2021;268:4178–4189. https://doi.org/10.1007/s00415-021-10512-x
3. Walsh, AJ, Ghosh, A, Brain, AO, Buchel, O, Burger, D, Thomas, S, White, L, Collins, GS, Keshav, S, Travis, SPL. Comparing disease activity indices in ulcerative colitis, J Crohn's Colitis. 2014;8(4):318–325. https://doi.org/10.1016/j.crohns.2013.09.010
4. Jandhyala R. Neutral theory: applicability and neutrality of using current clinical trial endpoints where disease severity scores are available. 2021. [In peer review].
5. Jandhyala, R. Neutral Theory: A conceptual framework for construct measurement in clinical research. 2021. [Preprint].
6. Jansen-van der Weide, MC, Gaasterland, CMW, Roes, KCB, et al. Rare disease registries: potential applications towards impact on development of new drug treatments. Orphanet J Rare Dis. 2018;13(1):154. doi:10.1186/s13023-018-0836-0
7. Heise, W, Kersten, O, Kassner, KM, Birkenmeyer, G, Grosse, G, Niedobitek, F. Fulminant primary manifestation of Crohn's colitis "Hot Crohn's disease". Z Gastroenterol. 1997 Jun;35(6):481–90. PMID: 9231992.
8. Vermeire, S, Schreiber, S, Sandborn, WJ, Dubois, C, Rutgeerts, P. Correlation Between the Crohn's Disease Activity and Harvey–Bradshaw Indices in Assessing Crohn's Disease Severity. Clin Gastroenterol Hepatol. 2010;8(4):357–363, ISSN 1542-3565, https://doi.org/10.1016/j.cgh.2010.01.001
9. Cortina-Borja M, te Vruchte D, Mengel E, et al. Annual severity increment score as a tool for stratifying patients with Niemann-Pick disease type C and for recruitment to clinical trials. Orphanet J Rare Dis. 2018;13(1):143. doi:10.1186/s13023-018-0880-9
10. Jandhyala R. Neutral theory: applicability and neutrality of using generic health-related quality of life tools in diseases or conditions where specific tools are available. BMC Med Res Methodol. 2021;21(1):86. doi:10.1186/s12874-021-01279-w

11. Ropars, J, Gravot, F, Ben Salem, D, Rousseau, F, Brochard, S, Pons, C. Muscle MRI. A biomarker of disease severity in Duchenne muscular dystrophy? A systematic review. Neurology. 2020;94(3):117–133; doi: 10.1212/WNL.0000000000008811
12. Ouzzani, M., Hammady, H., Fedorowicz, Z. et al. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5:210. https://doi.org/10.1186/s13643-016-0384-4
13. Reilly FE. Charles Peirce's theory of scientific method. Fordham Univ Press; 2018.
14. Liszka JJ. Peirce's convergence theory of truth redux. Cognitio: Revista de Filosofia. 2019;20(1):91–112. doi:10.23925/2316-5278.2019v20i1p91-112
15. Flanagan RF, Dammann O. The epistemological weight of randomized-controlled trials depends on their results. Perspect Biol Med. 2018;61(2):157–173. doi:10.1353/pbm.2018.0034
16. Lim, J-A, Lee S-T, Moon, J et al. Development of the clinical assessment scale in autoimmune encephalitis. Ann Neurol. 2019;85(3):352–358. https://doi.org/10.1002/ana.25421
17. Cai MT, Lai QL, Zheng Y, et al. Validation of the Clinical Assessment Scale for Autoimmune Encephalitis: A multicenter study. Neurol Ther. 2021;10(2):985–1000. doi:10.1007/s40120-021-00278-9
18. Rossetti AO, Logroscino G, Milligan TA, Michaelides C, Ruffieux C, Bromfield EB. Status Epilepticus Severity Score (STESS). J Neurol. 2008;255(10):1561–1566. doi:10.1007/s00415-008-0989-1
19. Gao Q, Ou-Yang T, Sun X, Yang F, Wu C, Kang T, et al. Prediction of functional outcome in patients with convulsive status epilepticus: The END-IT score. Crit Care. 2016;20:46
20. Leitinger, M., Höller, Y., Kalss, G. et al. Epidemiology-Based Mortality Score in Status Epilepticus (EMSE). Neurocrit Care. 2015;22:273–282. https://doi.org/10.1007/s12028-014-0080-y
21. Mittal MK, Rabinstein AA, Hocker SE, Pittock SJ, M Wijdicks EF, McKeon A. Autoimmune encephalitis in the ICU: analysis of phenotypes, serologic findings, and outcomes. Neurocrit Care. 2016;24(2):240–250. doi:10.1007/s12028-015-0196-8
22. Tao R, Qin C, Chen M, et al. Unilateral cerebral cortical encephalitis with epilepsy: a possible special phenotype of MOG antibody-associated disorders. Int J Neurosci. 2020;130(11):1161–1165. doi:10.1080/00207454.2020.1720676
23. Pilotto A, Masciocchi S, Volonghi I, et al. The clinical spectrum of encephalitis in COVID-19 disease: the ENCOVID multicentre study. 2020:2020.06.19.20133991. doi:10.1101/2020.06.19.20133991
24. National Institute for Health and Care Excellence. Betibeglogene autotemcel for treating transfusion-dependent beta-thalassaemia. Appraisal consultation document. 2021.
25. ClinicalTrials.gov. Efficacy and safety of Lucerastat oral monotherapy in adult subjects with Fabry disease (MODIFY). Identifier: NCT03425539. 2021. Available from: https://clinicaltrials.gov/ct2/show/NCT03425539?cond=lucerastat&draw=2&rank=3

Figures

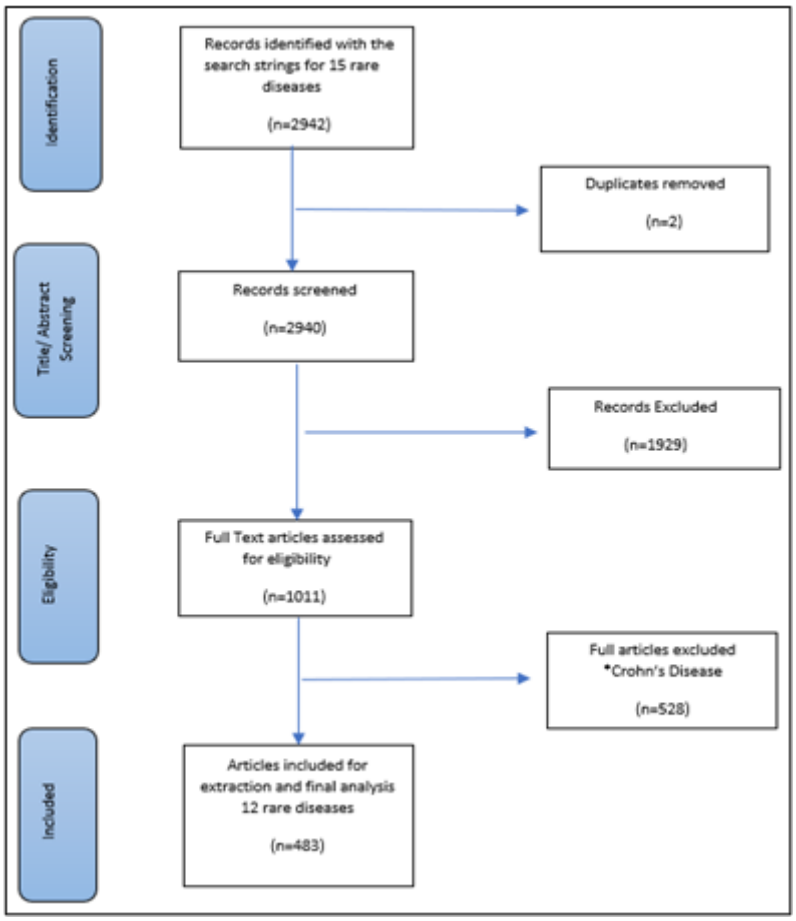

Figure 1

Selection of articles for use to assess the neutrality of clinical studies

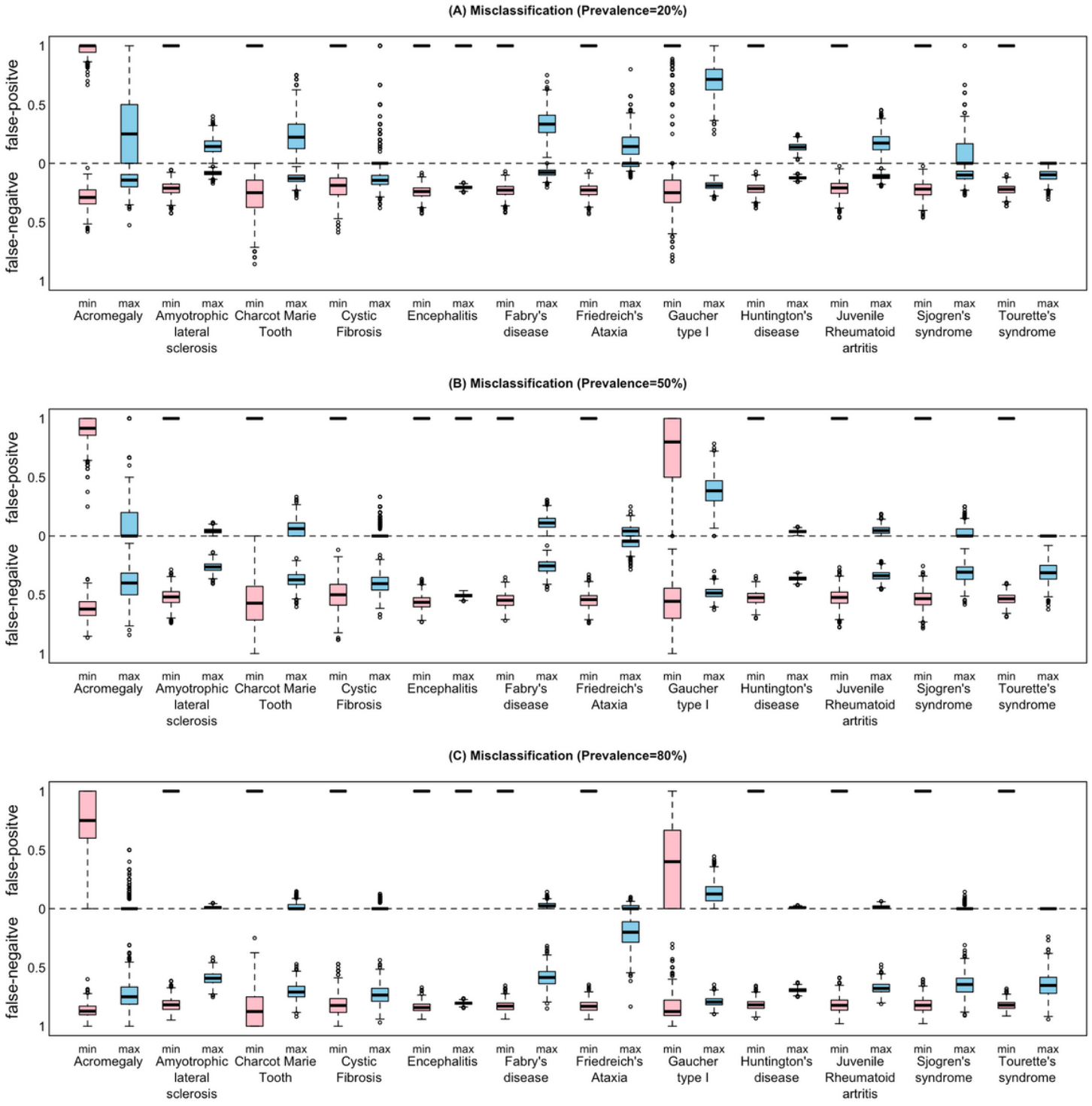

Figure 2

Potential misclassifications for all diseases in respect of the first DSS

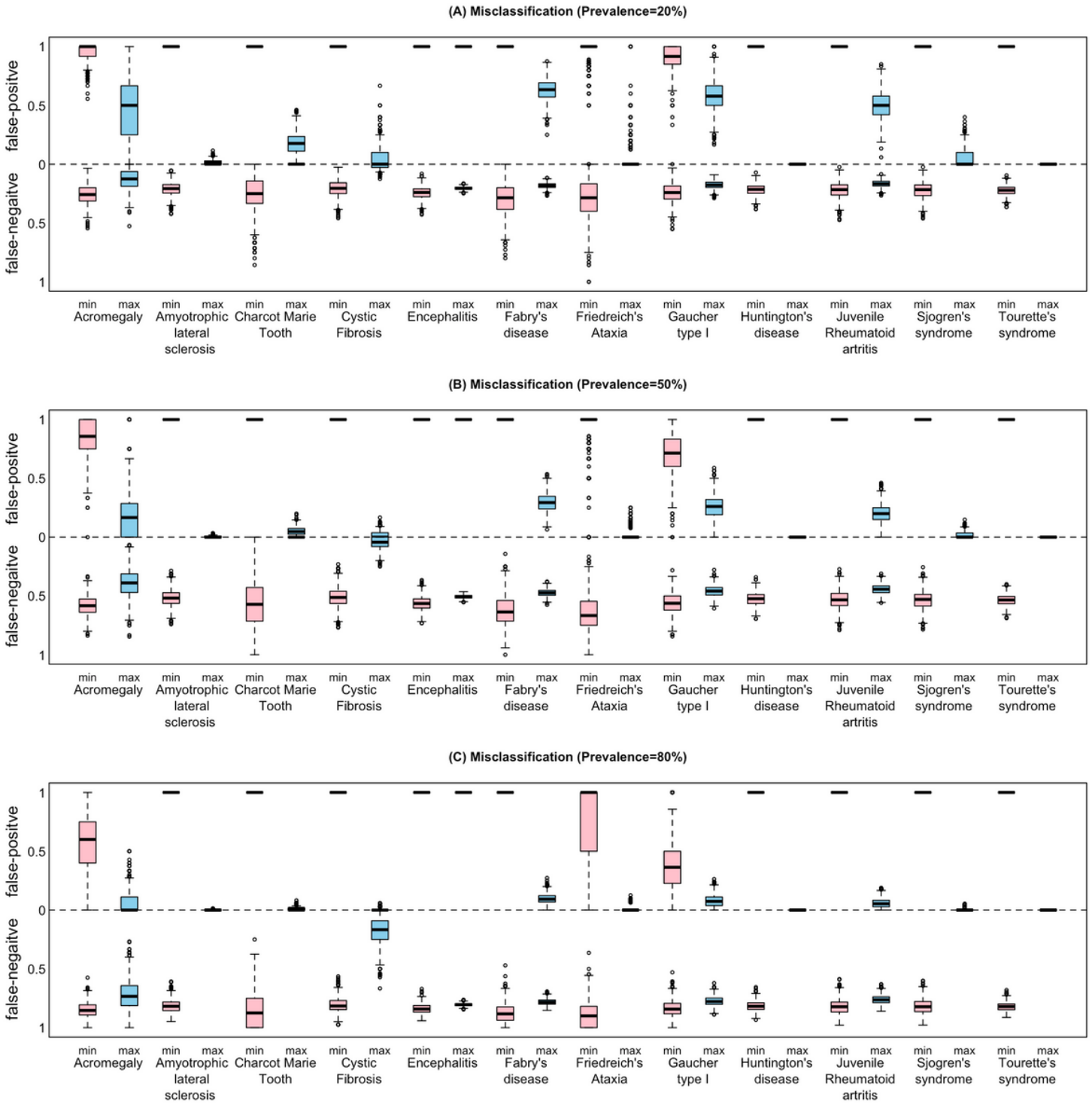

Figure 3

Potential misclassifications for all diseases in respect of the composite DSS