The Accuracy Of Network Meta-Analysis Feasibility Predictions Based On Data Included Only In The Abstracts Of Trial Publications

Rutherford L†, Ahdesmäki O, Clarke N, Martin A, Witkowski M
Crystallise, Stanford-le-Hope, Essex, United Kingdom

OBJECTIVES: Network meta-analysis (NMA) is an effective technique for comparing multiple treatments simultaneously but requires treatment regimens and outcomes to be reported consistently across trials. Time-consuming data extraction from full texts is needed to confirm whether a network can be created. We propose a time-saving method that uses data from publication abstracts to determine whether an NMA might be feasible. **METHODS**: A recent systematic literature review (SLR) with NMA comparing efficacy and safety of two antiretroviral therapies for HIV was used as the gold standard. Using an Evidence Mapper tool (www.evidencemapper.co.uk), abstracts from the 206 studies included in the SLR were indexed by fields including each comparison, trial name and reported outcomes. A network-feasibility tool assessed the possibility of creating a connected network including efavirenz and dolutegravir. The accuracy of the abstract-predicted networks for each outcome were compared to those in the published SLR. **RESULTS**: Viral suppression at 48 and 96 weeks, CD4 cell count at 48 and 96 weeks, discontinuations and treatment-related adverse event rates were the most commonly reported outcomes in both the full texts and abstracts. The abstract-only sensitivity for predicting networks for these outcomes was 92% to 100%. The accuracy of less commonly reported outcomes was lower, with overall sensitivities of 25% to 100% and specificities of 33% to 100%. Specificity was 100% for 15 of the 23 outcomes. Time taken from identifying relevant publications to network prediction for the abstract-only method was 98 hours compared to an estimated 600 hours with full data extraction. **CONCLUSIONS**: Although this method cannot confirm that an NMA is feasible, it could create a network of sufficient accuracy for common outcomes to identify when an NMA cannot be undertaken. When NMA is feasible, subsequent data extraction can be quicker by focusing on those interventions and outcomes that can be compared.

Introduction

Randomised controlled trials (RCTs) are considered the gold standard for assessing the efficacy and safety of health interventions and are often placed at the top of evidence hierarchies. However, single RCTs rarely provide sufficient information for addressing the demands of patients, clinicians and policymakers. Instead, each trial provides a piece of evidence that, when taken together with others provides a standard basis for evidence-based healthcare decision making. [1] SLRs are a useful tool for combining evidence from individual RCTs, consolidating the vast amounts of research on a specific topic and making comparisons of all relevant competing interventions. [2] In general, when multiple RCTs compare a subset of interventions of interest, it is possible to develop a network of RCTs where all trials have at least one intervention in common with another. [3] This network of interlinked RCTs allows for both direct and indirect comparisons of interventions by means of a NMA. [4–7]

NMAs offer a set of methods to combine quantitative results of comparable studies. They enable readers to visualise and interpret the wider picture of evidence and to understand the relative merits of multiple interventions where individual RCTs have not investigated head to head comparisons. [2,3,6] Even when the results of direct comparisons are conclusive, combining them with the results of indirect comparisons may yield more precise findings as a greater evidence base is considered. [3,6,8] Since the body of primary literature has increased exponentially over the past 20 years, conducting SLRs and NMAs has become a necessity to more accurately and comprehensively present scientific knowledge. [9] NMAs are increasingly being performed to inform decision-making regarding the comparative efficacy and safety of alternative healthcare treatments. [10]

A vital assumption of NMA is that studies within a network are consistent in terms of variables that may

† Corresponding Author
E-mail address: contact@crystallise.com

impact the magnitude of treatment effects. [4,5,11] It is important to confirm that the underlying evidence base informing treatment comparisons consists of studies similarly conducted, with similar patient populations and characteristics, and similar outcome measures. [5,12–14] Sound healthcare decision making requires comparisons of all relevant competing trials. Ideally, robustly designed RCTs would simultaneously compare all interventions of interest. However, such studies are almost never available. [3,15–17] In this case, it is necessary to assess the feasibility of performing a valid NMA.

Conducting an NMA feasibility assessment involves several steps designed to evaluate whether differences among studies may affect treatment comparisons or make some comparisons invalid. [4] The first step to plan and document the process for a feasibility assessment is to establish the research question of interest through development of a project protocol. The second step involves identifying and 'mapping' the relevant evidence from clinical trial data and establishing intervention comparisons (direct and indirect networks) and outcome availability across relevant studies. One of the best ways to locate this key evidence is through a recently published SLR. The third step requires an assessment of heterogeneity within the network. At this stage it is useful to inspect patient enrollment criteria across trials, compare the distribution of effect modifiers such as baseline demographic data, and consider the importance of variations between study methodology for trial conduct and outcome measure. The final step involves collating all information from step one through three to consider the best approaches for NMA. [18] Depending on the available data, an NMA may be feasible. However, if there is no connected network containing the most critical indirect comparisons of interest, alternative approaches should be considered. These may include creating a list of assumptions, performing a meta-regression analysis, subgroup analysis, sensitivity analysis or summarising why an NMA is not feasible. [19]

Unfortunately, given the ever-growing numbers of studies published, NMA feasibility assessments can require large amounts of time and effort to search the literature and summarise findings, often with uncertainty in time and resources required for completion. [9,20,21] For example, a recent meta-analysis of SLRs registered in the International Prospective Register of Systematic Reviews (PROSPERO) registry to quantify the time and resources required to complete such projects estimated the time to project completion and publication to be 67.3 weeks (IQR = 42). The number of studies identified through literature searches ranged from 27 to a staggering 92,020, with a mean yield rate of 2.94% (IQR = 2.5) of included studies. Additionally, the mean number of authors was five per review (SD = 3). This highlights the sheer volume of time and resources required in the NMA feasibility process. [20]

To address this problem, researchers have proposed several streamlining strategies for SLRs. Growing evidence suggests that expedite methods, such as rapid reviews, can be conducted with relatively minor implications on validity. [22–25] However, trading off scientific rigor for speed when creating a knowledge basis is controversial, and the consequences are insufficiently known. [22,25–29] The current study proposes a novel time saving method that uses data from publication abstracts rather than full texts to determine whether an NMA might be feasible.

Aim

This study presents a method that employs data extracted solely from publication abstracts to conduct an initial NMA feasibility assessment. This methodology was applied within the framework of a recent SLR with NMA focusing on the comparative efficacy and safety of antiretroviral therapies for HIV. The primary objective is to enhance the initial assessment of the potential suitability of NMA for an early stage of systematic review process. This may be used to predict the existence of connected networks and streamline the subsequent analytical processes.

Methodology

Data Source and Study Selection

The primary data source for this study consisted of publications included in a recent SLR with NMA. A database search was carried out to identify SLRs with NMAs on topics of interest. The identified SLRs were screened to find publications that adhered to the following criteria:

- Free full-text publication available for accessibility and transparency.
- Full list of publications included in SLR and NMA.
- NMA involving >10 interventions and >5 outcomes with sufficient level of detail reported.
- SLR/NMA with subgroup analyses.

A recent SLR with NMA publication focusing on the comparative efficacy and safety of antiretroviral therapies for HIV was selected. [30] The data used to

create the network in this study had come from full-texts and so could be used as the gold standard when the accuracy of the abstract-predicted and full-text methods were compared.

The principal analysis of this SLR focused on treatment-naïve adults and adolescents, but it comprehensively addressed a range of sub-populations, including tuberculosis (TB) co-infected patients, pregnant and breastfeeding women, patients with pre-treatment drug resistance (PDR), women and adolescents of childbearing age, and individuals with prior exposure to antiretroviral drugs (ARVs). While the broad inclusion criteria for the primary population were designed to capture these sub-populations, no evidence base was identified.

As a result, all publications included in the SLR formed the basis of our research and were indexed to facilitate the initial feasibility assessment based on the title and abstract content. Our analysis focused on studies included in the primary analysis described in the published SLR with NMA.

The full-text publication with supplementary materials of the selected SLR with NMA were downloaded. A complete list of included studies was extracted and compiled with available bibliographic details (doi, authors, title, abstract, full citation, URL). This dataset was exported into a csv file and then uploaded into the Evidence Mapper, a web-based evidence mapping software (www.evidencemapper.co.uk).

Indexing and Fields

Fields for Early NMA Feasibility Assessment

The selected indexing fields were specifically chosen to facilitate our initial assessment of network meta-analysis feasibility. These fields are as follows:

- **Disease**: the type of HIV if stated e.g., HIV-1.
- **Study Type**: RCT, non-randomised comparative study or non-comparative study.
- **Subpopulations**: the characteristics of the population being studied e.g., co-infection with tuberculosis or hepatitis, pregnant or breastfeeding women, adults, children or infants.
- **Year**: Recording the publication year to contextualize the research chronology.
- **Location**: Specifying the geographic locations where the studies were conducted.
- **Study Size**: the number of patients included in the study.
- **Study Duration**: Noting the reported study duration, which provides temporal context.
- **Risk Factors**: any risk factors for being infected with HIV e.g., fetal exposure.
- **Outcomes**: the specific outcomes reported with their timeline when given in the abstract e.g., viral load at week 96.

Fields Detailing Assessed Intervention Regimens

Complete HAART Regimen with Doses: drugs used in the highly active antiretroviral therapy (HAART) regimen with doses (total daily) included when available from the abstract. If more than one regimen was described, these were listed as separate tags.

Drugs within the HAART Regimen Being Compared: the purpose of this indexing field was to highlight which drugs were being compared between the multi-drug regimens. Each comparison was listed as a separate tag.

Comparison of Complete HAART Regimens with Doses: for the comparative studies, the comparison between the highly active antiretroviral therapy (HAART) regimen for each study was listed as a separate tag e.g., regimen A vs B, regimen A vs C, regimen B vs C. Doses (total daily) were included when available from the abstract. For the majority of tags, the full names of the drugs have been used. However, due to a character limit some tags could not be created for some of the comparisons. In these cases, the drugs' abbreviations were used.

Fields Tailored for Networker Tool Usage

The following fields were specifically selected for the Networker tool and indexed according to following rules:

- **Full Citation**: This field serves as a unique identifier and holds the full bibliographic citation, including author details, publication year, title, journal information, and digital identifiers such as DOI and URL. It plays a pivotal role in data processing within the Networker tool.
- **Trial Name**: The name of the trial or if this was not stated, the NCT number or first author plus year of publication e.g., Albano_2019. This field serves as the source for network diagram edges, minimizing the risk of trial overcounting by merging records with similar Trial Names and Interventions. To maintain data integrity and accuracy, the Networker tool combines multiple records with the same values for Trial Name and Interventions fields into a single record, thereby consolidating outcomes reported across all abstracts. Conversely, in instances where

an abstract reports results for multiple trials, the Networker tool separates records by Trial ID, creating distinct records for each trial while retaining the original column data.

- **Third Agent Drugs Only**: This field comprises a list of unique interventions compared within the trials, serving as the source for nodes in network diagram generation. The node definitions align with the approach outlined by Kantars et al. (2022), [30] emphasizing the characterization of nodes in terms of specific antiviral drugs rather than full Antiretroviral Therapy (ART) regimens. Notable exceptions include differentiating between multiple doses of efavirenz (600 mg qd and 400 mg qd).
- **Heterogeneity Assessment**: This field is used for grouping based on the initial heterogeneity assessment, with tags that indicate inclusion in the primary analysis or within specific subgroups or subpopulation groups. It is used in filtering data within the Networker tool, enhancing precision and relevance.

Early NMA Heterogeneity Assessment

Data Fields

In our early NMA heterogeneity assessment, we examined the studies based on available data in fields extracted from publication abstracts, including disease, study type, subpopulations, study duration, risk factors, and outcomes. These fields were pivotal in determining the heterogeneity across identified studies and making decision on including studies to the primary analysis or other groups.

Our heterogeneity assessment criteria were aligned with methods used by Kanters et al., [30] where primary analysis included treatment-naïve adults and adolescents HIV patients.

The results were indexed in the "Heterogeneity Assessment" field, with tags indicating inclusion in the primary analysis or within specific subgroups or subpopulation groups. This field was used for filtering data within the Networker tool, thereby enhancing precision and relevance in our feasibility assessment.

Outcome Comparison and Sensitivity-Specificity Analysis

Data on interventions contributing to outcome-specific networks were collected from the supplement tables in the SLR (Table 11 to Table 38). The results were synthesized and summarized in a table, wherein each outcome was represented in rows and interventions in columns.

The next step involved the export of the evidence map indexing generated from abstract data using the Evidence Mapper tool. The evidence map data were imported into the Networker tool, which facilitated the creation of network diagrams for each outcome under investigation, as well as summary tables on outcome-specific networks (number of trials and list of interventions contributing to each outcome analysis). Subsequently, we compared the results obtained from the Networker tool to the outcomes reported in the source SLR.

To evaluate the agreement between our abstract-based indexing and trials incorporated in the NMA for each outcome, we computed the "Ratio." This metric signifies the ratio of trials identified in our analysis to trials encompassed in the NMA for each specific outcome. A ratio exceeding 1 signifies that our analysis identified a greater number of trials, whereas a ratio below 1 implies that the NMA encompassed more trials.

For two outcomes encompassing multiple follow-up durations (CD4 count and Viral suppression), we aggregated the findings for each outcome into a higher-level grouping to assess the concordance between our abstract-based indexing and NMA. This approach aimed to evaluate how our methodology could predict a network diagram on a broader scale, without taking into account specific follow-up durations.

To quantify the congruence between the SLR and our abstract-based mapping, we employed sensitivity and specificity analyses for each outcome with following criteria:

- **True Positive (TP)**: The interventions identified by our evidence mapping and present in the outcome-specific NMA.
- **True Negative (TN)**: The interventions not present in our evidence mapping and not present in the outcome-specific NMA.
- **False Positive (FP)**: The interventions identified by our evidence mapping but not present in the outcome-specific NMA.
- **False Negative (FN)**: The interventions not identified by our evidence mapping but present in the outcome specific NMA.
- **Sensitivity** was calculated as: TP/(TP+FN)
- **Specificity** was calculated as: TN/(TN+FP)

Estimation of time taken for standard SLR-NMA process compared to evidence mapping approach

In order to explore the potential time-saving element of this process using data extracted from abstracts only, we estimated and then compared the time taken to make an NMA feasibility decision using a standard SLR_NMA process with the evidence mapping approach. The estimated time taken was calculated for the following processes: searching, title and abstract screening, shortlisting, full text screening and data extraction. For the more focused abstract-only data extraction we used the time taken to index an abstract for all relevant fields when creating an evidence map. These estimations of time were calculated using our internal costing sheets for evidence synthesis projects.

Results

In total, 151 publications describing 68 studies were included in the primary analysis. Thirty-three outcomes were considered for the NMA in the gold standard SLR. Out of these, twelve outcomes were excluded from our analysis. Five of these were excluded due to the NMA not being performed because of limited/low quality data; the remaining seven were excluded because the data available in the abstracts were limited or conflicting and so these outcomes were not indexed in our Evidence Map (Table 1).

In total, thirteen interventions were considered for inclusion in the NMA (see Table 4 in appendix for list of interventions and their abbreviations).

Table 1. Outcomes

Outcome	NMA	Our analysis
Any neuropsychiatric AE	Yes	Multiple tags grouped (n=21)*
CD4 count at 24 weeks	Yes	Yes
CD4 count at 48 weeks	Yes	Yes
CD4 count at 96 weeks	Yes	Yes
CD4 count at 144 weeks	Yes	Yes
Discontinuations	Yes	Multiple tags grouped (n=8)*
Discontinuations due to AEs	Yes	Yes
Dizziness any grade	Yes	Yes
Emergent AEs	Yes	Multiple tags grouped (n=6)*
Overall resistance	Yes	Yes
Sleep disorders any grade	Yes	Multiple tags grouped (n=5)*
Treatment related AEs	Yes	Yes
Treatment related SAEs	Yes	Yes
Viral suppression at 4 weeks	Yes	Yes
Viral suppression at 12 weeks	Yes	Yes
Viral suppression at 24 weeks	Yes	Yes
Viral suppression at 48 weeks	Yes	Yes
Viral suppression at 96 weeks	Yes	Yes
Viral suppression at 144 weeks	Yes	Yes
Weight change at 48 weeks	Yes	Yes
Weight change at 96 weeks	Yes	Yes
Development of AIDS-defining illnesses	Yes	Outcome not reported in abstract
Dizziness grade 3 or 4	Yes	
Emergent SAEs	Yes	
NRTI resistance	Yes	
Suicidal ideation	Yes	
Third resistance	Yes	
Treatment-emergent NNRTI resistance	Yes	
Depression any grade	No NMA	Excluded from our analysis as no NMA was conducted for these outcomes, due to no or low quality data.
Depression grade 3 or 4	No NMA	
Mortality	No NMA	
Viral suppression among >100k at 48 weeks	No NMA	
Viral suppression among >100k at 96 weeks	No NMA	

AE: adverse events; SAE: serious adverse events; NRTI: nucleotide reverse transcriptase inhibitors; NNRTI: non-nucleoside reverse transcriptase inhibitors; *See Tables 5, 6, 7 and 8 in the appendix for outcomes included in the multiple (or umbrella) tag groups

Outcome Analysis and Sensitivity-Specificity Assessment

Viral suppression at 48 and 96 weeks, CD4 cell count at 48 and 96 weeks, discontinuations, and treatment-related adverse event rates were the most commonly reported outcomes in both the full texts and abstracts.

In our comparative outcome analysis between the SLR and our abstract-based indexing, we observed varying degree of concordance in the number of trials reporting outcomes (Table 2).

- Two outcomes (any neuropsychiatric, Emergent adverse events (AEs)): ratio >1 (number of trials reporting these outcomes higher in our abstract-based indexing, compared to SLR/NMA)
- Eight outcomes (CD4 at 24, 48 and 144 weeks; Viral suppression at 4, 12, 24, and 144 weeks; weight change at 96 weeks): ratio <0.5 (number of trials

reporting these outcomes notably lower in our abstract-based indexing, compared to SLR/NMA)

Eleven outcomes (remaining): ratio ≥0.5 to 1.0 (number of trials reporting these outcomes lower in our abstract-based indexing or same, compared to SLR/NMA).

Table 2. Outcome Analysis Summary

Outcome	Number of studies reporting outcome		
	SLR	EM	Ratio
Any neuropsychiatric AE	7	23	3.29
Discontinuations	58	30	0.52
Discontinuations due to AEs	56	52	0.93
Dizziness any grade	17	10	0.59
Emergent AEs	32	45	1.41
Overall resistance	35	23	0.66
Sleep disorders any grade	15	9	0.60
Treatment related AEs	31	25	0.81
Treatment related SAEs	18	18	1.00
Weight change at 48 weeks	2	2	1.00
Weight change at 96 weeks	5	2	0.40
CD4 count at 24 weeks	28	3	0.11
CD4 count at 48 weeks	43	12	0.28
CD4 count at 96 weeks	25	15	0.60
CD4 at 144 weeks	7	2	0.29
CD4 (follow-ups combined)	50	28	0.56
Viral suppression at 4 weeks	33	5	0.15
Viral suppression at 12 weeks	31	3	0.10
Viral suppression at 24 weeks	46	13	0.28
Viral suppression at 48 weeks	58	46	0.79
Viral suppression at 96 weeks	35	28	0.80
Viral suppression at 144 weeks	12	4	0.33
Viral suppression (follow-ups combined)	67	57	0.85

AE: adverse events; SAEs: serious adverse events; Outcomes in bold denote most commonly reported

The abstract-only sensitivity for predicting networks for the most commonly reported outcomes was 92% to 100%. The accuracy of less commonly reported outcomes was lower, with overall sensitivities of 25% to 100% and specificities of 33% to 100%. Specificity was 100% for 15 of the 21 outcomes. The results of sensitivity and specificity assessment are presented in Table 3.

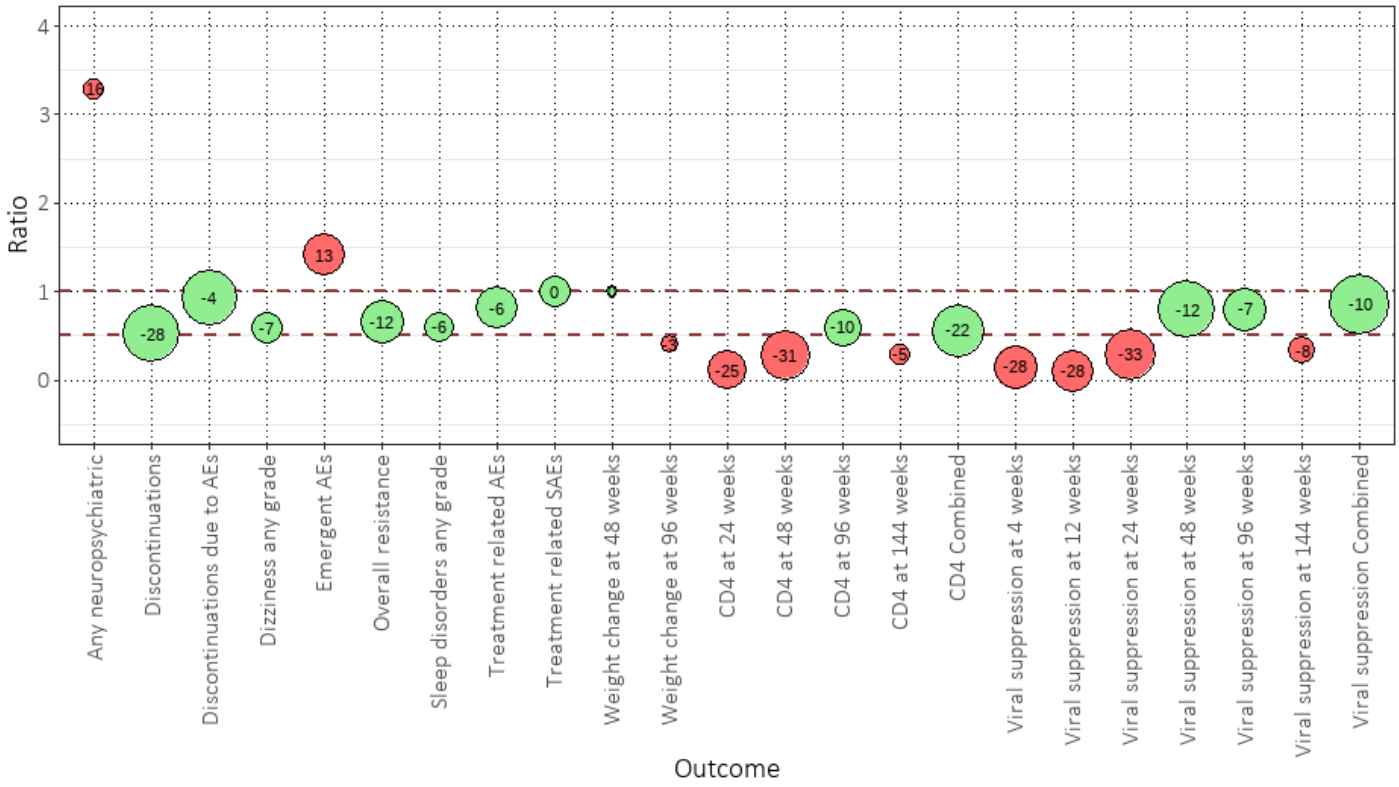
High sensitivity scores were observed for outcomes such as 'Discontinuations,' 'Emergent AEs,' 'Overall resistance,' and 'Treatment related SAEs,' each demonstrating a perfect sensitivity of 100%. Conversely, outcomes with notably lower sensitivity scores, included CD4 at 24 (45%) and 144 (29%) weeks, Viral suppression at 4 (50%), 12 (25%), and 144 weeks (50%), and Weight change at 96 weeks (50%). For the remaining outcomes, a range of sensitivity values was observed, with scores ranging from 63% to 92%, signifying varying degrees of accuracy in identifying relevant trials.

Table 3 Sensitivity and Specificity Assessment Results

Outcome	Sensitivity	Specificity
Any neuropsychiatric	100%	40%
Discontinuations	**100%**	**100%**
Discontinuations due to AEs	100%	50%
Dizziness any grade	75%	100%
Emergent AEs	100%	100%
Overall resistance	83%	100%
Sleep disorders any grade	63%	80%
Treatment related AEs	**92%**	**100%**
Treatment related SAEs	100%	100%
Weight change at 48 weeks	67%	80%
Weight change at 96 weeks	50%	86%
CD4 count at 24 weeks	45%	100%
CD4 count at 48 weeks	**83%**	**100%**
CD4 count at 96 weeks	**73%**	**50%**
CD4 count at 144 weeks	29%	100%
CD4 (follow-ups combined)	92%	100%
Viral suppression at 4 weeks	50%	100%
Viral suppression at 12 weeks	25%	100%
Viral suppression at 24 weeks	67%	100%
Viral suppression at 48 weeks	**92%**	**-***
Viral suppression at 96 weeks	**92%**	**100%**
Viral suppression at 144 weeks	50%	100%
Viral suppression (follow-ups combined)	92%	-*

* NMA for viral suppression at 48 weeks included all interventions, therefore specificity is not calculated (no TP values); AE: adverse events; SAEs: serious adverse events; outcomes in bold denote most commonly reported

Figure 1 Outcome Analysis



Number in circle express the difference in number of trials identified in our evidence map and number of trials included in the Kanters et al. NMA. Positive value indicate that our indexing identified more trials reporting specific outcome, compared to the SLR.

The difference in sensitivities and outcome analysis ratios at the various follow-up times for CD4 cell count and viral suppression is noteworthy. The higher sensitivity values for the two outcomes at week 48 (83% and 92%) and 96 (73% and 92%) are as a result of these being the follow-up times points of most interest (and therefore priority was given to including these in the abstracts).

The lower sensitivity scores for extremes of the follow-up times (24 weeks and 144 weeks for CD4 cell count) and (4 weeks and 144 weeks for viral suppression) reflect that although they were reported in the full text publications, they were not the main follow-up times of interest and so were not included in the abstracts.

It is a similar observation for the outcome analysis ratios. The ratios for CD4 cell count at 24 and 144 weeks and viral suppression at 4 weeks and 144 weeks are lower (0.11 and 0.29; 0.15 and 0.33) than at 96 weeks (0.6) for CD4 cell count and 48 weeks and 96 weeks for viral suppression (0.79 and 0.8) meaning that the number of trials reporting these outcomes is notably lower in our abstract-based indexing compared to the SLR/NMA because they are considered a less useful outcome for this particular NMA. Interestingly, the ratio for CD4 count at 48 weeks is also low at 0.28

The sensitivities for the outcomes 'Discontinuations' and 'Discontinuations due to AEs' are also worth comment. Poor adherence to antiretroviral therapy can increase the likelihood of developing resistance therefore measuring adherence via discontinuation in general or due to more specific reasons would have been an important outcome for the researchers to measure. Again, being a priority outcome would have meant it more likely to be included in the abstract and resulting in a high sensitivity score in our analysis.

While the sensitivity scores for both discontinuation outcomes are 100%, the outcome analysis ratios differ (0.52 for discontinuations; 0.93 for discontinuations due to AEs) giving us insight into how the discontinuations due to AEs are of particular importance and therefore reported in the abstract. The same principle applies to the outcomes 'Treatment related AEs' and 'Treatment related serious AEs'. These are important outcomes to measure as they affect the patient's level of adherence and quality of life so are more likely to be included in the abstract. The

heat map of interventions against outcomes using the sensitivity/specificity analysis criteria shows us that for the majority of interventions, most outcomes measured were identified in the abstract as well as being present in the NMA (Figure 2). Notable exceptions include Doravirine where all outcomes were not reported in the abstract or full text apart from viral suppression at 48 weeks which was reported in the full text but not the abstract.

If the heat map is read by outcome, then similar observations to the outcome analysis ratio and sensitivity/specificity analysis are seen. For the CD4 cell count at 24 and 144 weeks, viral suppression at 4, 12, 24 and 144 weeks, there are more false negatives meaning that the outcome was in the full text but not the abstract, and more true negatives meaning that the outcome was not reported in either.

The results for the combined groups, which aggregate outcomes across various follow-up durations, demonstrate notable patterns. For the combined groups (CD4 count and viral suppression) our method exhibited a high sensitivity of 92% and a specificity of 100%.

Comparison of time taken

The number of hours taken to reach the stage where an NMA feasibility decision could be made was approximately 98 hours using the Evidence Mapping approach and approximately 600 hours using the standard SLR approach. Figure 2 shows how the two methods differ in time taken and process.

Discussion

Our study showed that it is possible to use an approach that employs data extracted only from trial abstracts to correctly predict the existence of connected networks and therefore conduct an initial NMA feasibility assessment. This is evidenced by the most commonly reported outcomes identified by our method aligning with the primary outcomes used in the NMA suggesting that for the studies included in the NMA, the most relevant outcomes were reported in the abstracts.

To assess the agreement between our abstract-based indexing and trials incorporated in the NMA for each outcome, we calculated an outcome analysis ratio. This ratio compared the number of trials reporting outcomes in the full-text data extraction method and our abstract-based indexing method. Over half of the outcomes, which included the primary ones, had a ratio of between 0.5 and 1 which signifies that the same or slightly lower number of trials reported outcomes in both methods. To calculate the accuracy of abstract-based indexing in identifying relevant studies compared to the full-text data extraction we carried out a sensitivity and specificity analyses for each outcome. The analysis demonstrated that the abstract-only sensitivity for predicting networks for the most commonly reported outcomes was high with results between 92% and 100%. This high level of sensitivity shows us that the relevant outcomes reported in the full text publications were also reported in the abstracts. Furthermore, fourteen of the outcomes had a specificity of 100% showing that our method was not identifying data that was not deemed relevant to this particular NMA.

Figure 2. Heat map of the outcomes against interventions using the sensitivity/specificity analysis criteria

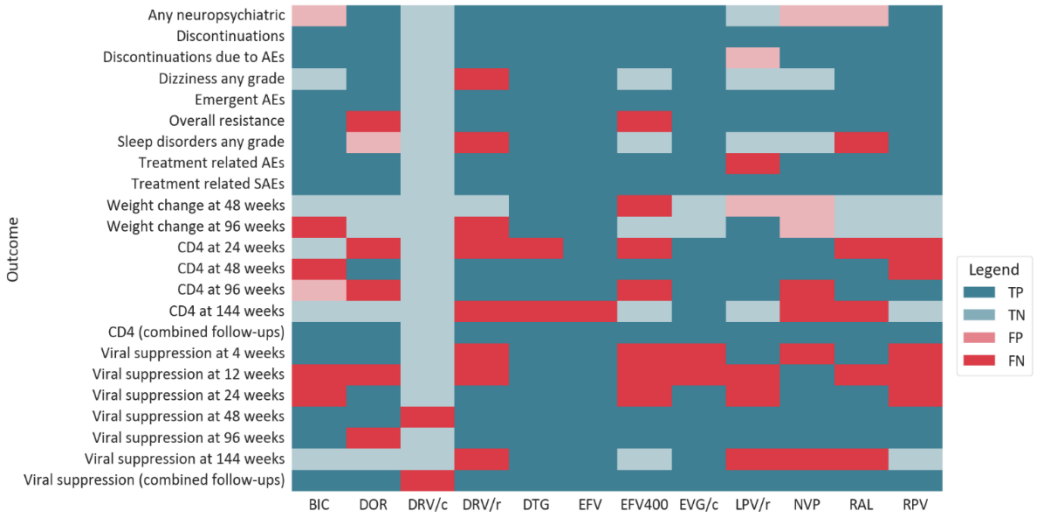
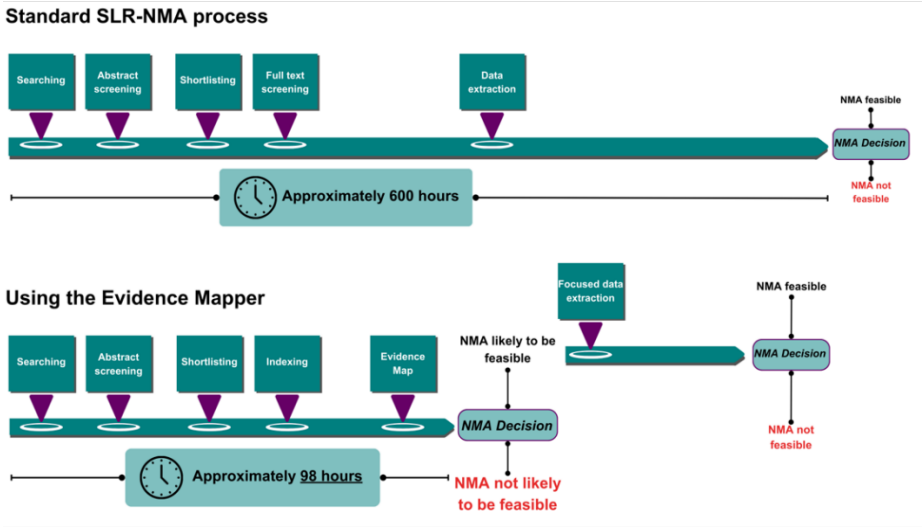


Figure 3. Infographic showing comparison of time taken to make a decision about NMA feasibility



We also demonstrated that the time saved by using an abstract-based method was approximately six-fold. A large number of resources are used in the standard SLR full-text data extraction to the point where it is thought there is sufficient information for an NMA feasibility decision to be made. We have shown that full-text data extraction may not be required to make an early NMA feasibility decision as there can be sufficient data in the abstracts.

Limitations

Although we aligned the abstract indexing to the detail provided in the NMA as closely as possible, there were limitations to our method. For word count reasons, there is inevitably less detail provided in the abstracts compared to the full texts. We found that although the outcomes of interest were usually reported, specific details of the outcome were sometimes missing. For example, the time points at which the CD4 count and viral suppression were collected were occasionally not available. This was reflected in the notably lower sensitivity scores and for CD4 cell count at 24 and 144 weeks (45% and 29%) whereas the sensitivity scores and for the main follow-up times of 48 and 96 weeks are much higher (83% and 73%;). The abstracts could give details of the complex HIV drug regimens in an abbreviated format. Indexing errors might have occurred if the abbreviated format was not interpreted correctly.

We observed that two outcomes, 'Any neuropsychiatric AE' and 'Emergent AEs' were associated with outcome analysis ratios >1 meaning that the number of trials reporting these outcomes was higher in our abstract-based indexing compared to the SLR/NMA. Interestingly for these outcomes we had used a functionality of the Evidence Mapping tool that can group related outcomes together under one 'umbrella tag'. This can be useful in organising the indexed data but for this study, and for this particular NMA, it resulted in us over-counting outcomes, especially for the neuropsychiatric adverse events which included 21 different tags. However, in other early feasibility studies for a different NMA, being able to group outcomes together could be useful.

One further limitation was that unlike Kanters et al., [30] who conducted detailed full-text data analysis, our study relied solely on data available in abstracts. Consequently, studies included by Kanters et al. [30] in their SLR but excluded from the analysis set remained in our dataset. This leads to overinclusion, characterized by the incorporation of a greater number of studies and can influence the interpretation of the early NMA feasibility assessment. On one hand, it may suggest the existence of specific connections (edges) within the network, potentially expanding the scope of the analysis. However, in a worst-case scenario, overinclusion may erroneously indicate a connected network when, in reality, it remains disconnected (following detailed NMA feasibility assessment).

Conclusions

This method provided an efficient and precise way to identify suitable studies and assess the feasibility of constructing a network, saving valuable research time and enhancing the rigor of the subsequent network meta-analysis.

References

1. IntHout J, Ioannidis JP, Borm GF. Obtaining evidence by a single well-powered trial or several modestly powered trials. Stat Methods Med Res. 2016 Apr;25(2):538–52.

2. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2. [Internet]. The Cochrane Collaboration; 2009. Available from: http://www.cochranehandbook.org.

3. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005 Oct 15;331(7521):897–900.

4. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting Indirect-Treatment-Comparison and Network-Meta-Analysis Studies: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 2. Value Health. 2011 Jun;14(4):429–37.

5. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. Value Health. 2011 Jun;14(4):417–28.

6. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct 30;23(20):3105–24.

7. Salanti G, Higgins JP, Ades A, Ioannidis JP. Evaluation of networks of randomized trials. Stat Methods Med Res. 2008 Jun;17(3):279–301.

8. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. Stat Med. 2003 Oct 15;22(19):2995–3016.

9. Bastian H, Glasziou P, Chalmers I. Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up? PLoS Med. 2010 Sep 21;7(9):e1000326.

10. Bafeta A, Trinquart L, Seror R, Ravaud P. Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review. BMJ. 2013 Jul 1;347(jul01 1):f3675–f3675.

11. 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;50(6):683–91.

12. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med. 2009 Jun 30;28(14):1861–81.

13. 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.

14. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC Med. 2013 Dec;11(1):159.

15. Ioannidis JP. Indirect comparisons: the mesh and mess of clinical trials. The Lancet. 2006 Oct;368(9546):1470–2.

16. Sutton A, Ades AE, Cooper N, Abrams K. Use of Indirect and Mixed Treatment Comparisons for Technology Assessment: PharmacoEconomics. 2008;26(9):753–67.

17. Wells G, Sultan S, Chen L, Khan M, Coyle D. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. [Internet]. Canadian Agency for Drugs and Technologies in Health; 2009. Available from: https://www.cadth.ca/sites/default/files/pdf/H0462_itc_tr_e.pdf

18. Hutton B. Feasibility Assessments for NMA: An Underutilized Tool [Internet]. 2023. Available from: https://www.linkedin.com/pulse/feasibility-assessments-nma-underutilized-tool-brian-hutton/

19. Cope S, Zhang J, Saletan S, Smiechowski B, Jansen JP, Schmid P. A process for assessing the feasibility of a network meta-analysis: a case study of everolimus in combination with hormonal therapy versus chemotherapy for advanced breast cancer. BMC Med. 2014 Dec;12(1):93.

20. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic

reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017 Feb;7(2):e012545.

21. Mills EJ, Thorlund K, Ioannidis JPA. Demystifying trial networks and network meta-analysis. BMJ. 2013 May 14;346(may14 2):f2914–f2914.

22. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. Implement Sci. 2010 Dec;5(1):56.

23. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev. 2012 Dec;1(1):10.

24. Khangura S, Polisena J, Clifford TJ, Farrah K, Kamel C. RAPID REVIEW: AN EMERGING APPROACH TO EVIDENCE SYNTHESIS IN HEALTH TECHNOLOGY ASSESSMENT. Int J Technol Assess Health Care. 2014 Jan;30(1):20–7.

25. Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, et al. Rapid reviews versus full systematic reviews: An inventory of current methods and practice in health technology assessment. Int J Technol Assess Health Care. 2008 Apr;24(02):133–9.

26. Harker J, Kleijnen J. What is a rapid review? A methodological exploration of rapid reviews in Health Technology Assessments. Int J Evid Based Healthc. 2012 Dec;10(4):397–410.

27. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. J Clin Epidemiol. 2012 Dec;65(12):1267–73.

28. Sagliocca L, De Masi S, Ferrigno L, Mele A, Traversa G. A pragmatic strategy for the review of clinical evidence. J Eval Clin Pract. 2013 Aug;19(4):689–96.

29. Thomas J, Newman M, Oliver S. Rapid evidence assessments of research to inform social policy: taking stock and moving forward. Evid Policy. 2013 Jan;9(1):5–27.

30. Kanters S, Vitoria M, Zoratti M, Doherty M, Penazzato M, Rangaraj A, et al. Comparative efficacy, tolerability and safety of dolutegravir and efavirenz 400mg among antiretroviral therapies for first-line HIV treatment: A systematic literature review and network meta-analysis. EClinicalMedicine. 2020 Nov;28:100573.

Crystallise Whitepaper: The Accuracy Of Network Meta-Analysis Feasibility Predictions Based On Data Included Only In The Abstracts Of Trial Publications

Appendix

Appendix Table 1 Interventions included in the NMA and their abbreviations

Intervention	Abbreviation
Bictegravir	BIC
Doravirine	DOR
Darunavir boosted with cobicistat	DRV/c
Darunavir boosted with ritonavir	DRV/r
Dolutegravir	DTG
Efavirenz	EFV
Efavirenz (400 mg)	EFV400
Elvitegravir	EVG/c
Lopinavir	LPV/r
Nevirapine	NVP
Raltegravir	RAL
Rilpivirine	RPV

Appendix Table 2 Tags included in the 'Any neuropsychiatric adverse event ' umbrella tag

Altered sensorium	Neurocognitive impairment
Anxiety	Neurocognitive performance
Asthenia	Neurologic AEs
Brain volume changes	Neurological defects
Disturbances in attention	Neuropathy
Drug related CNS AEs	Neuropsychiatric AEs
Drug related psychiatric AEs	Paranoia
Dysesthesia	Pre-specified CNS events
Fatigue	Somnolence
Hallucinations	Vertigo
Headache	

Appendix Table 3 Tags included in the 'Discontinuations' umbrella tag

Completion rate	Drug switching
Discontinuations	Efficacy related treatment discontinuation
Discontinuations due to lack of efficacy	Lost to follow up
Drug discontinuations	Renal-related discontinuations

Appendix Table 4 Tags included in the 'Emergent adverse events' umbrella tag

Adverse events	Patient reported symptoms
Adverse events in infants	Tolerability
Grade 2 to 4 AEs	Toxicities

Appendix Table 5 Tags included in the 'Sleep disorders any grade' umbrella tag

Abnormal dreams	Sleep disorder
Insomnia	Sleep disturbances
Nightmares	