3-D evidence network plot system: a new software tool to aid the exploration of covariate imbalances and effects in network meta-analysis

Sarah Batson^{1*}, Robert Score¹ and Alex J Sutton²

¹DRG Abacus, 6 Talisman Business Centre, Talisman Road, Bicester, UK, OX26 6HR

²Department of Health Sciences, University of Leicester, Centre for Medicine, University Road, Leicester, UK, LE1 7RH

*Corresponding author

Telephone: +44 (0) 1869 241281

Email: SBatson@teamDRG.com

Abstract

Objective

To develop the 3-dimensional (3-D) evidence network plot system - a novel web-based interactive 3-D tool to facilitate the visualisation and exploration of covariate distributions and imbalances across evidence networks for network meta-analysis (NMA).

Study design and setting

We developed the 3-D evidence network plot system within an AngularJS environment using a third party JavaScript library (Three.js) to create the 3-D element of the application. Data used to enable the creation of the 3-D element for a particular topic are inputted via a Microsoft® Excel template spreadsheet that has been specifically formatted to hold these data. We display and discuss the findings of applying the tool to two NMA examples considering multiple covariates. These two examples have been previously identified as having potentially important covariate effects, and allow us to document the various features of the tool while illustrating how it can be used.

Results

The 3-D evidence network plot system provides an immediate, intuitive, and accessible way to assess the similarity and differences between the values of covariates for individual studies within and between each treatment contrast in an evidence network. In this way, differences between the studies, which may invalidate the usual assumptions of a NMA, can be identified for further scrutiny. Hence, the tool facilitates NMA feasibility/validity assessments and aids in the interpretation of NMA results.

Conclusion

The 3-D evidence network plot system is the first tool designed specifically to visualise covariate distributions and imbalances across evidence networks in 3-D. This will be of primary interest to

systematic review and meta-analysis researchers, and, more generally, those assessing the validity and robustness of an NMA in order to inform reimbursement decisions.

Key words

Evidence networks, meta-analysis feasibility, network meta-analysis, novel graphical tool, 3-dimensional

Running title

3-D evidence networks

Word count

4,997

What is new?

- Visually assimilating, exploring and interpreting the distribution of covariate values across trials in a network meta-analysis (NMA) is challenging due to the complexities of representing the network structure simultaneously alongside study-level covariate values.
- This article describes a 3-dimensional (3-D) evidence network plot system-a novel, freely-accessible, web-based package to facilitate the exploration of covariate distributions and imbalances across evidence networks in NMA.
- The primary innovation which allows for the extensions to evidence networks and improvements is the use of a 3-D graphical environment, incorporating the graphical representation of covariates on a third 'z' axis.
- We believe this work to be the first application of a 3-D graphical environment to evidence networks in NMA.
- We propose that the 3-D evidence network plot system will facilitate the exploration of covariate distributions and imbalances across evidence networks and be of most value in the context of supporting NMA feasibility/validity assessments and to aid in the interpretation of NMA results to a wide audience.

1. Introduction

Network meta-analysis (NMA) is an increasingly popular statistical method used for estimating the comparative efficacy of all treatments of interest for a given condition, by simultaneously synthesising data from all relevant randomised controlled trials (1). Such analyses are commonly used to identify the most effective treatments and inform economic decision models to estimate the relative cost-effectiveness of the treatment options.

Like all statistical modelling, NMA makes a number of assumptions that, if not satisfied by the data being synthesised, can lead to erroneous results and misleading conclusions (2). The first assumption that needs satisfying is that the network is connected which can be checked by constructing a network diagram (3). More generally, evidence network diagrams are commonly used for visualising the available evidence base for the purpose of assessing the feasibility of the meta-analysis and understanding the strength and diversity of the evidence available. A conventional network diagram consists of 'nodes' representing the treatments of interest and edges representing available direct comparisons between pairs of interventions and are a key component of global NMA reporting checklists (4). All nodes should be connected to form a single network via edges and any nodes which are not connected should be excluded. The amount of available evidence can also be presented in network diagrams by 'weighting' the nodes and edges using different node sizes and line thicknesses (5).

Further assumptions of NMA relate to the comparability of the studies being combined. As for pairwise meta-analysis, differences in the results of studies (beyond that expected by chance) within each (pairwise) treatment comparison is described as between-study heterogeneity. Such variability in study results can lead to inconsistency in treatment estimates across different comparisons in an NMA, where estimates of comparative effectiveness differ between those from direct comparisons and those derived from indirect comparison routes through the network (6). While heterogeneity and inconsistency random-effect (RE) terms can be included in NMA models to allow for them (7), results can become increasingly difficult to interpret as the number and magnitude of such terms increase. This can lead to challenging issues, in terms of limiting the ability to generalise from the results (8), for both decision-makers (9) and for designers of further studies that are intended to update the evidence base in the future (10, 11). Therefore, it is highly desirable to explain the causes and magnitude of heterogeneity and inconsistency rather than simply accommodate them.

Heterogeneity and inconsistency can frequently be explained by the differences in trial design and the conduct of the individual trials included in the NMA. Assuming summary information from published trial results are being used for the NMA rather than individual patient data (12), it may be possible to identify causes of heterogeneity and inconsistency by extracting information on study- and aggregate patient-level characteristics (e.g. duration of treatment or duration of condition prior to randomisation). Such variables are often described as potential effect modifiers and if these impact the effectiveness of the interventions of interest, treatment by covariate interactions can be included in the NMA model (7). Treatment-covariate interactions can be used to explain and reduce heterogeneity and inconsistency in the same way as they are used in meta-regression for pairwise meta-analysis (13). In addition, when treatment by covariate interactions relating to patient characteristics are identified, it implies that treatment efficacy varies between patients. Therefore, optimal treatment decisions could vary across patient groups depending on their characteristics. In

the current paper we focus on potential effect modifiers which are expressed on a continuous scale (including dichotomous patient-level covariates aggregated at the study level, e.g. % male), although we note that categorical variables (e.g. individual indicators of study quality) can also be considered using a regression framework and plots including these have been considered elsewhere (14). Regression modelling is generally superior to subgroup analyses as it allows a holistic analysis, exploring the impact of covariates on all of the data, and allows the simultaneous consideration of multiple (continuous and categorical) covariates (8). However, it should not be forgotten that regressing study-level summary covariate information on study-level average treatment effects is potentially susceptible to ecological bias.

A recent publication (14), outlining a process for assessing the feasibility of conducting a valid NMA, highlighted the importance of assessing whether there are differences in treatment, patient and outcome characteristics across comparisons that may affect the summary measures of treatment effects relative to an overall reference treatment. These potential effect modifiers may be known or suspected a priori, or identified post hoc. Visually assimilating, exploring and interpreting the distribution of covariate values across trials in an NMA is challenging due to the complexities of representing the network structure simultaneously alongside study-level covariate values. While multiple plots could more easily be constructed for individual comparisons within the network, these are of limited use because many will be sparse and uninformative, and each plot only provides a subset of the required information. A holistic approach is required to assess the distribution of covariate values across the whole evidence network.

Cope et al. (14) present what we believe to be the first published attempt to present such information graphically for categorical and aggregate patient-level dichotomous covariates (i.e. expressed as a %). This was achieved by including pie-charts for each study close to the relevant comparison edge on a network diagram (see Cope et al. Figures 4 and 5 for examples). While we acknowledge the utility of such an approach, we find assimilating the information from such displays challenging, particularly as it is not always clear which comparisons certain study pie-charts relate to. Furthermore, no method of displaying information from continuous covariates is proposed.

In this paper, we present a novel, interactive, freely-accessible, web-based package to present continuous and aggregate patient-level dichotomous covariate information superimposed on an evidence network. We see our work as complementary to that previously discussed (14), but we believe our method is easier to interpret for aggregate patient-level dichotomous covariate information (the context in which both approaches could be used) as it circumvents the issues highlighted above. The primary innovation underlying the suggested improvements is the use of an interactive 3-dimensional (3-D) graphical environment. Such 3-D models are easier to interpret when they can be rotated rather than the reliance on static images. The readers of this publication are encouraged to explore the tool in conjunction with this paper (available at <http://3dnma.com/plot>).

In the current paper Section 2 outlines the rationale for and features of the software tool we have developed; Section 3 introduces the first illustrative NMA example from rheumatoid arthritis (RA), which has a relatively simple star-shaped evidence network and is presented as an introduction to the concept of the 3-D network plot system using disease duration and baseline risk as two continuous covariates of interest (8); Section 4 introduces a second example in advanced breast cancer, taken from the paper by Cope et al. (14). Here we graphically present one of the continuous

covariates considered by Cope et al. (not originally presented graphically) and an aggregate dichotomous patient-level covariate-visceral metastases. The latter was presented visually by Cope et al. and we report a comparison with our 3-D network plot system. Section 5 introduces advanced applications of the system whereby we demonstrate the reduction of network data to a single summary covariate per treatment contrast and incorporate data relating to the precision of the outcome data.

2. Features of the 3-D network plot system

The 3-D network plot system allows the incorporation of a third 'z' axis to display covariate bars for each trial included within a network. The covariate bars for each study are placed on the edges of the relevant comparisons and projected on the third dimension.

First, a dataset is uploaded to the system in the form of a Microsoft® Excel sheet; a template can be downloaded from the system. When a covariate is not selected, the system produces a conventional two-dimensional (2-D) network diagram. This allows for examination of the network structure, instantaneous removal of treatment nodes from the dataset, and for conducting sensitivity analyses or producing focused views on a proportion of a complex network. When a covariate is selected, the system incorporates a third z-axis and 'bars' to represent the covariate measure for each trial. This produces a 3-D representation which allows for the distribution of covariate effects for each comparison to be simultaneously displayed. The system can hold multiple covariates and each covariate can be selected from a drop-down menu. The system also features zoom functionality and an automatic rotation function (which can be used simultaneously with other manual controls) to facilitate the exploration of data.

Following the initial input of data, many aspects can be manually configured. This includes the position of the treatment nodes on the x-y axis and the colouring of the nodes; the principal view can be centered on any treatment node of interest. The treatment nodes and treatment contrasts can be weighted to reflect the volume of available evidence. The covariate name, value, scaling, and units can be amended, in addition to the width of the covariate bars. The distance and positioning of the covariate bars across the treatment contrasts can also be modified and the colour can be changed to represent positive and negative values (relative to a reference value). Finally, the background grid feature and covariate labelling elements can be toggled on and off. All manual changes within the system can be saved and reloaded (as 'raw data') and the system is able to generate figures in Portable Network Graphics (PNG) format.

The software can be accessed at <http://3dnma.com/plot> and requires a user to register with their email address. The datasets for the examples used within this publication can be selected and loaded from within the system via a drop-down menu (example datasets 1-6). A full instruction manual is also available and can be downloaded from within the system.

3. Example 1: rheumatoid arthritis

The first example dataset includes trials from a review of the NICE technology appraisal for certolizumab pegol (CZP) for the treatment of RA in patients who had failed disease-modifying anti-

rheumatic drugs (DMARDs) (15). The dataset includes twelve methotrexate (MTX) controlled trials comparing seven different treatments (placebo + MTX, adalimumab + MTX, CZP + MTX, etanercept + MTX, infliximab + MTX, rituximab + MTX and tocilizumab + MTX), forming a 'star' shaped network where all comparisons are relative to placebo.

The dataset was used as an example in the NICE guidance on methods for meta-regression for the outcome of American College of Rheumatology 50% (ACR50) at 6 months (8). Two potential sources of effect modification are explored in the NICE guidance – disease duration and the baseline risk of ACR50. Baseline risk is defined as the risk of ACR50 in the placebo arm of each trial.

3.1. Disease duration

Screenshots from the covariate visualiser of the evidence network and an evidence network incorporating information on the average disease duration from each study are shown in Fig.1 and Fig.2. A 2-D view from the covariate visualiser (the x-y plane with no third z-axis displayed) is shown in Fig.1. and essentially represents a standard 2-D network plot; the grid function is not displayed in this example. Fig.2 shows the inclusion of the third (z) axis and the addition of covariate 'bars'-the grid is also added on this example to aid the interpretation of the image in 3-D. A single bar on the z-axis represents the covariate level of a single trial on the relevant edges connecting the treatment nodes compared in that trial. For example, the RAPID 1 and RAPID 2 trials compared placebo + MTX with CZP + MTX. Therefore, the two bars representing each of these trials lie on the edge connecting the placebo + MTX node with CZP + MTX. The disease duration for each trial has been centered on the mean disease duration across all trials of the network (8.21 years or 98.5 months). Consequently, in this example, the values on the z-axis may be positive or negative. Negative covariate values (representing trials in which disease duration is lower than the mean across all trials of the network) are presented as red bars, and positive values (representing trials in which disease duration is higher than the mean across the network) are presented as green bars. This further aids the interpretation of the distribution of the covariates across the network.

Disease duration was identified as a potential source of effect modification (8). Previous studies in RA have demonstrated that shorter disease duration is associated with smaller treatment effects compared with a longer disease duration (16).

Fig.1. Evidence network in 2-D

Fig 2. Evidence network displaying the disease duration covariate values for each trial

In Fig.2 it is evident that there is considerable variability in the mean disease duration of patients enrolled in the randomised controlled trials (RCTs) included in the network. The studies with the most extreme values are immediately identifiable as Weinblatt 1999 (high value) and the CHARISMA trial (low value). Importantly, it can be clearly seen that this variability is systematic. For example, the RAPID 1 & 2 trials, which compared CZP + MTX with placebo + MTX, were both conducted in patients with shorter disease duration compared with the mean disease duration of the network. Thus, in this network, there is likely to be an unfavourable bias against CZP + MTX as a shorter disease duration is associated with smaller treatment effects (16). A similar scenario can also be observed for tocilizumab + MTX. The single trial (Weinblatt 1999) which compares etanercept + MTX with placebo + MTX was conducted in patients with the longest disease duration of all the trials in the network. Thus, there is likely to be a favourable bias for etanercept + MTX in this network compared with other DMARDs as longer disease duration is associated with greater treatment effects (8, 16).

While the meta-regression models presented by Dias et al. were not strongly supported by the trial data, the estimate of the interaction coefficient from the RE model did not include the null value within its 95% credible interval (CrI) [0.14 (95% CrI 0.01, 0.26)] (8). Thus, consistent with the findings from larger studies in RA, the interaction coefficient from the RE adjusted model suggests that for each 1- year unit increase in disease duration, the log odds ratio (OR) of each comparator versus placebo + MTX increases by 0.14 (i.e. larger treatment effects are associated with longer disease duration). Therefore, in this example, the implications of disease duration interaction with treatment effects should be considered for the decision model.

3.2. Baseline risk

As the meta-regression models adjusting for disease duration were not strongly supported by the RA dataset from the technology appraisal, Dias et al. recommended that further explanations of the causes of heterogeneity should be sought (8). Baseline risk is the underlying risk of the outcome of interest within a study population and represents a summary of both known and unknown risk factors. Baseline risk is a potentially important source of heterogeneity, particularly among studies where the baseline risk varies.

The screenshots from the dataset visualised in the covariate visualiser are shown in Fig.3 and Fig.4. Fig.3 demonstrates the view along the x-z plane and Fig.4 incorporates all three axes. Together, these figures highlight considerable variability in baseline risk across the trials of the network. Note, this particular visualisation is only possible when all trials contain a common comparator treatment arm, such as that seen in 'star' shaped networks. In this example, two trials (Abe 2006 and CHARISMA) exhibit higher baseline risks compared with the other trials of the network. This could be a potential cause for concern highlighting unfavourable bias towards the active treatments of these trials.

The baseline risk models explored by Dias et al. using the RA dataset suggested strong interaction effects between the baseline risk of ACR50 and relative treatment effects; this relationship should be incorporated in cost effectiveness analyses (8). The estimate of the interaction coefficient from the RE model suggested that for each one unit (the difference from 100–0%) increase in the baseline log-odds of ACR50, the log OR of each comparator versus placebo + MTX decreased by 0.95 (i.e. the greater the ACR50 response in the placebo + MTX arm the smaller the treatment effect).

Fig.3. Evidence network displaying the baseline risk of placebo + MTX arm of each trial-view of the network along the x-z plane

Fig.4 Evidence network displaying the baseline risk of placebo + MTX arm of each trial

4. Example 2: advanced breast cancer

The second dataset is the case-study presented by Cope et al. which allows for the comparison of everolimus in combination with hormonal therapy with alternative therapies in terms of progression free survival (PFS) in women with advanced breast cancer (14). The dataset consists of 24 RCTs which form a connected evidence network for the PFS outcome (which include hormonal and chemotherapy comparators). Data for several potential effect modifiers for each trial within the network are tabulated in Cope et al. (14). This approach is consistent with the common format of reporting potential treatment effect modifiers in NMA publications which can be difficult to interpret. In addition, Cope et al. visually presented two of the aggregate dichotomous patient-level covariates (hormone receptor status and visceral metastases) as a pie-chart for each study, close to the relevant comparison edge on a network diagram (see Cope et al. Figures 4 and 5).

Cope et al. reported that differences within and across the treatment comparisons were observed in terms of estrogen receptor positive status, exposure to prior hormonal therapies, exposure to prior chemotherapies, and visceral metastases (14). However, Cope et al. also reported that differences in performance status and age were less prominent, while differences in post-menopausal status, human epidermal growth factor 2 status and types of hormonal therapies were noted (14). In the following sections we present one of the continuous covariates considered by Cope et al. (14) and an aggregate dichotomous patient-level covariate – visceral metastases. The latter was presented visually by Cope et al. (14) and we contrast our approach with theirs.

4.1. Continuous covariate-median age

We have explored all the potential treatment effect modifiers presented in Cope et al. (14) as 3-D network plots. Indeed, as suggested by Cope et al. (14) the differences in terms of age across the trials appear to be less prominent than those observed with the other potential treatment effect modifiers. On display of the median age covariate within the 3-D network plot system, it is immediately evident that there is relatively little variability in median age across the trials of the evidence network; in this example the median age values across the trials have not been centered to a reference value in the 3-D network plot system and therefore all bars are represented in green as the default colour for positive values (Fig.5). However, it is easy to identify the treatment comparisons associated with the most extreme median age values (Fig.5). For example, the median age of patients in the O'Shaughnessy 2001 trial (bar has been coloured red to highlight the trial in Fig.5) is 70 years. This is the highest across the network and is 10 years greater than the mean of median ages (60 years) across the trials of the network. The O'Shaughnessy 2001 trial is the only study that contributes to the direct comparison of cyclophosphamide + MTX + fluorouracil (CMF) versus capecitabine (CAP). This could lead to a potential bias in comparisons of CAP with other treatments of interest depending on the potential direction of effect modification by age. The median age of patients in the Ingle 1982 trial (bar has been coloured yellow to highlight the trial in Fig.5) is 49 years, which is the lowest across the network and is 11 years below the mean of median ages. However, Ingle 1982 is one of three trials contributing evidence for the direct comparison of tamoxifen (TAM) with megestrol acetate (MA). While it would be challenging to detect any covariate effect(s) of age with such little variability across the majority of studies, conducting a sensitivity analysis removing individual trials for which covariate values are potentially a concern, is possible.

Fig.5: Evidence network of the Cope et al. dataset displaying the median age covariate

4.2. Aggregate dichotomous patient level covariate-visceral metastases

Cope et al. reported visceral metastases data (present, absent or not reported) by presenting a network diagram together with pie-charts for each study next to the relevant comparison edge (see Cope et al. Figure 5) (14). We have visualised the visceral metastases covariate data via the 3-D network plot system as the percentage of patients with the outcome. Where a proportion of patients were reported to have visceral metastases, the remaining proportion of patients had no visceral metastases. Furthermore, any trial not reporting the covariate of interest is clearly highlighted with a black circle on the 2-D axis and labelled “N/R”. As displayed in Fig.6 and Fig.7, it is evident that there is variability in the proportion of patients with visceral metastases across the trials of the evidence network. The treatment comparisons that the trials with the most extreme values contribute to are also easily distinguishable (Fig.6 and Fig.7). For example, the trials contributing to the comparisons of docetaxel (DOC) with nab-paclitaxel (Nab-PAC), liposomal doxorubicin (L-Dox) or vinorelbine (VIN) enrolled patients with the highest proportion of visceral metastases (87–91%). In comparison, all trials which included a treatment of TAM or MA included less than 50% of patients with visceral metastases.

Fig.6: Evidence network of the Cope et al. dataset displaying the proportion of patients with visceral metastases-view 1

Fig.7. Evidence network of the Cope et al. dataset displaying the proportion of patients with visceral metastases-view 2

5. Advanced application and features of the 3-D network plot system

5.1. Reduction of data to a single covariate bar per treatment contrast

As an alternative to presenting covariate values for each trial, the system can be used to present an overall summary of the average covariate value across all trials for each treatment contrast. This allows for an overall summary of the level of the covariates for each treatment comparison. The exact influence a trial's covariate value has on estimation of network meta-analysis parameters varies by parameter and thus is difficult to represent by a single value (17). Therefore we suggest examining both the weighted and unweighted mean covariate value (weighting can be based on $1/\text{variance}$ of each trial's treatment effect). We have presented this summary for the covariate of disease duration in RA (example 3 dataset). Here, the weighted mean of the covariate values (centered to the network weighted mean) is presented for all studies in a given treatment comparison (Fig.8). Fig.8 can be compared with Fig.2 to facilitate the comparison between presenting summary information per treatment contrast versus all trial level evidence per contrast, respectively (although an unweighted mean was used to centre the covariate values in Fig.2). Note that there are three trials for the treatment contrast of placebo + MTX compared with adalimumab + MTX, two of which have a higher disease duration than the mean of the network, while a single trial has a lower disease duration than the mean of the network (Fig.2). However, for the treatment contrast of placebo + MTX compared with adalimumab + MTX, the summary of the covariate for this contrast is higher than the weighted mean of the network.

Fig.8. Evidence network displaying the weighted mean level of disease duration covariate per treatment contrast

5.2. Incorporation of data relating to the precision of outcome data

The width of the covariate bars are adjustable within the system. A good use for this functionality, to add more relevant information into the plot, as to further aid the interpretation of the data, is to make the bars proportional to the precision ($1/\text{variance}$) of the primary outcome for each of the trials. In this way the precise trials, providing most information and thus being most influential in the synthesis, are more visually dominant (in the same way as having the central estimate of effect plotting symbol being proportional to the precision of the study in a conventional forest plot). Here we have applied this approach to the RA example for the covariate of disease duration in a network

which presents all trial level covariate data (Fig.9) (example 4 dataset) and a network which presents summary data for each contrast (Fig.10) (example 5 dataset). These plots illustrate the influence of each of the trials within the networks. For example, Abe 2006 is one of the smallest trials of the network and it is clear from Fig.9 that it has a small influence on the relative efficacy of infliximab + MTX versus placebo + MTX compared with the two other trials contributing to this comparison (ATTEST and START trials). More generally, in Fig.9 and Fig.10, respectively, it is evident that the outcome data for the single trials contributing to the treatment contrasts of rituximab + MTX and etanercept + MTX versus placebo + MTX are associated with large variances (small precisions). Therefore, it is likely that the results from contrasts involving these treatment nodes will be associated with greater levels of uncertainty compared with other treatment nodes of the network.

Fig.9. Evidence network displaying the disease duration covariate values for each trial- covariate bar widths reflect the inverse variance weighting of each trial

Fig. 10. Evidence network displaying the weighted mean level of disease duration covariate per treatment contrast- covariate bar widths reflect the inverse variance weighting of the pooled data available for each treatment contrast

6. Discussion

We have developed a novel interactive tool which produces a 3-D version of a network diagram, incorporating the graphical representation of covariates on a third z-axis. The tool is available as an open access resource (available at <http://3dnma.com/plot>). We propose that this novel graphical approach can be used to facilitate the exploration of covariate distributions and imbalances across evidence networks. We believe it will be most valuable in the context of supporting NMA feasibility/validity assessments and aiding in the interpretation of NMA results.

In terms of NMA feasibility/validity assessments, the 3-D network plot system can support this by aiding in the identification of (i) imbalances in covariates considered to be potential effect modifiers and (ii) appropriate analyses to model and adjust for potential effect modifiers, identified as being imbalanced across treatment comparisons. Further, the tool can aid in the interpretation of NMA results through (i) constructing the plots for using data from published NMAs as an appraisal/assessment tool to check that important covariate imbalances do not exist, in the presence of heterogeneity, for covariates not adjusted for in the original analysis and (ii) reconciling the results of a meta-regression analysis through the impact of covariate adjustment versus unadjusted model results via the visualisation of covariate values across the whole network.

The proposed 3-D plotting system was designed to provide a high level of flexibility and functionality to the user, be that data driven or visual. The majority of the development was conducted within an AngularJS environment using a third party JavaScript library (Three.js) to create the 3-D element of the application. By using Three.js, it was possible to rapidly develop the application as Three.js supported a large proportion of the processing required for 3-D development. The 3-D element of the plot was developed using data held within a specially formatted Microsoft® Excel spreadsheet. The use of Microsoft® Excel to develop the system means that the application is both intuitive and user friendly. The system is also able to accommodate the manual input of multi-arm trials although the examples presented in this publication include two-arm trials only (see example 6 dataset from within the system).

In a pairwise meta-analysis context, exploratory plots of the data to explore potential associations between treatment effects and continuous study-level covariates are relatively straightforward. Such plots can also be overlaid with the results of a meta-regression analysis. Software is available to plot the effect size of individual studies against the covariate of interest (18). This software incorporates a plotting symbol, where the size of the symbol is proportional to the precision of the study. A fitted regression line of best fit can then be superimposed on top of this together with associated confidence and/or CrIs—an example of which can be found in Ibrahim et al. 2015, Figure 3 (19). Work is underway to develop an approach to incorporate *both* study outcome/ modelling results and study covariate values clearly for a NMA.

Future developments proposed for this application include the ability to produce stacked bar charts to visualise categorical covariates. In addition, integrating the system with an algorithm to optimally locate treatment nodes should help to make the network appear less cluttered (20). One interesting option would be to further develop the 3-D element of the system, such as allowing the system to be viewed in stereoscopic 3D (as found in virtual reality technologies) which may assist with interpretation (21). The ability to print the networks using a 3-D printer to produce real-life

sculptures to further aid in the understanding and interpretation of evidence networks is now a possibility. As digital media replaces paper, some journals are now starting to permit publication of 3-D figures and animations (22). We would fully support this initiative in applied statistics and in medical research generally.

7. Conclusion

The 3-D evidence network plot system is the first tool designed specifically to visualise covariate distributions and imbalances across evidence networks in 3-D. This will be of primary interest to systematic review and meta-analysis researchers, and, more generally, those assessing the validity and robustness of an NMA in order to inform reimbursement decisions. It will facilitate the exploration of covariate distributions and imbalances across evidence networks and be of most value in the context of supporting NMA feasibility/validity assessments and aid in the interpretation of NMA results. Given its free availability on the web, and simple spreadsheet interface it has the potential to make this powerful tool available to a wide audience.

Acknowledgements: None

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not for profit sectors.

Conflicts of interest: None

8. References

1. 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;347:f3675.
2. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. *Med Decis Making*. 2013;33(5):597-606.
3. 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;33(5):607-17.
4. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. 2015;162(11):777-84.
5. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PloS one*. 2013;8(10):e76654.
6. 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 in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2011;14(4):417-28.
7. 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;28(14):1861-81.
8. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Making*. 2013;33(5):618-40.
9. Ades AE, Lu G, Higgins JP. The interpretation of random-effects meta-analysis in decision models. *Med Decis Making*. 2005;25(6):646-54.
10. Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR. Evidence-based sample size calculations based upon updated meta-analysis. *Stat Med*. 2007;26(12):2479-500.
11. Welton N, Ades AE. Research decisions in the face of heterogeneity: what can a new study tell us? *Health Econ*. 2012;21(10):1196-200.
12. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221.
13. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559-73.
14. 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;12:93.
15. National Institute for Health and Clinical Excellence. Certolizumab pegol for the treatment of rheumatoid arthritis. TA186. 2010.
16. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med*. 2007;26(6):1237-54.
17. Donegan S, Dias S, Tudur Smith C, Marinho V, Welton N. Graphs of study contributions and covariate distributions for network meta-regression. *Res Synth Methods*. in press.
18. Harbord RM, Higgins JP. Meta-regression in Stata *The Stata Journal* 2008;8(4):493-519.
19. Ibrahimi P, Jashari F, Bajraktari G, Wester P, Henein MY. Ultrasound assessment of carotid plaque echogenicity response to statin therapy: a systematic review and meta-analysis. *Int J Mol Sci*. 2015;16(5):10734-47.

20. Rucker G, Schwarzer G. Automated drawing of network plots in network meta-analysis. *Res Synth Methods*. 2016;7(1):94-107.
21. Ware C, Franck G, editors. Viewing a graph in a virtual reality display is three times as good as a 2D diagram. *IEEE Conference on Visual Languages*; 1994 4-7 Oct 1994: IEEE.
22. Grossman T, Chevalier F, Kazi RH. Bringing research articles to life with animated figures. *Interactions*. 2016;23(4):52-7.