SPECIAL ISSUE PAPER

Visualizing the evolution of evidence: Cumulative network meta-analyses of new generation antidepressants in the last 40 years

Yan Luo[1] | Anna Chaimani[2] | Toshi A. Furukawa[1] | Yuki Kataoka[3] |
Yusuke Ogawa[4] | Andrea Cipriani[5] | Georgia Salanti[6]

[1]Department of Health Promotion and Human Behavior, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Research Center of Epidemiology and Statistics Sorbonne Paris Cité (CRESS-UMR1153), INSERM, INRA, Université de Paris, Paris, France

[3]Hospital Care Research Unit, Hyogo Prefectural Amagasaki General Medical Center, Hyogo, Japan

[4]Department of Healthcare Epidemiology, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan

[5]Department of Psychiatry, University of Oxford, Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK

[6]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

Correspondence
Toshi A. Furukawa, Department of Health Promotion and Human Behavior, School of Public Health in the Graduate School of Medicine, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.
Email: furukawa@kuhp.kyoto-u.ac.jp

It is often challenging to present the available evidence in a timely and comprehensible manner. We aimed to visualize the evolution of evidence about antidepressants for depression by conducting cumulative network meta-analyses (NMAs) and to examine whether it could have helped the selection of optimal drugs. We built a Shiny web application that performs and presents cumulative NMAs based on R netmeta. We used a comprehensive dataset of double-blind randomized controlled trials of 21 antidepressants in the acute treatment of major depression. The primary outcomes were efficacy (treatment response) and acceptability (all-cause discontinuation), and treatment effects were summarized via odds ratios. We evaluated the confidence in evidence using the CINeMA (Confidence in Network Meta-Analysis) framework for a series of consecutive NMAs. Users can change several conditions for the analysis, such as the period of synthesis, among the others. We present the league tables and two-dimensional plots that combine efficacy, acceptability and level of confidence in the evidence together, for NMAs conducted in 1990, 1995, 2000, 2005, 2010, and 2016. They reveal that through the past four decades, newly approved drugs often showed initially exaggerated results, which tended to diminish and stabilize after approximately a decade. Over the years, the drugs with relative superiority changed dramatically; but as the evidence network grew larger and better connected, the overall confidence improved. The Shiny app visualizes how evidence evolved over years,

emphasizing the need for a careful interpretation of relative effects between drugs, especially for the potentially amplified performance of newly approved drugs.

Highlights

- Network meta-analysis is considered to be a proper way of demonstrating the available evidence, since it allows comparisons between multiple interventions, and has been proved to be statistically powerful.
- It is challenging to present the voluminous results of NMA in an efficient and comprehendible manner.
- Evidence evolution based on the relatively new method NMA has not been investigated yet.
- The results of NMA should not only include the effects but also the confidence in the evidence, which can help interpret the findings appropriately.
- Effective use of rapidly developing statistical analysis and presentation tools such as Shiny package in R, may facilitate and simplify the visualization of NMA output.
- We should stay conservative towards new drugs, as their performance was often shown to be exaggerated initially, and it took time to become stable.

1 | INTRODUCTION

Modern medicine has reached a consensus that clinical practice should be guided by up-to-date evidence. However, the evidence changes rapidly, as many new agents and interventions are developed at a high speed and an exploding amount of clinical studies are published every day. It remains a challenge for investigators and practicing clinicians to appreciate the existing evidence appropriately in a timely manner. A tool that synthesizes and presents the often fragmented and ever-growing body of evidence would be of great value.

Evidence synthesis methods are developing fast. Pairwise meta-analyses can combine randomized controlled trials (RCTs) comparing two interventions. Network meta-analysis (NMA) enables the comprehensive and simultaneous comparison of multiple treatments. An empirical research has demonstrated that NMA, in comparison with pairwise meta-analysis, can increase statistical power and allow earlier detection of differences between treatment alternatives.[1] It is essential that the output from such an influential evidence resource like NMA is demonstrated in a succinct and understandable manner.

The evidence for treating patients with depression offers a typical example. Although antidepressants have been recommended as the first-line treatment, it has not been easy for physicians to choose drugs with better efficacy and safety among the many available. Early evidence, mostly based on RCTs and pairwise meta-analyses, implied that antidepressants had comparable efficacy. However, in the past 10 years, NMAs suggested that different antidepressants might have clinically important differences in efficacy and acceptability.[2,3] It is however unclear whether this evidence base, if available earlier, would have been able to indicate differences between the drugs and how evidence has evolved with additional trials.

Although NMA can help answer questions like this, it is challenging to present its output, since it usually produces a multitude of information and findings, from network diagrams to pairwise treatment comparisons and tests for the underlying assumptions. Additionally, statements about the credibility of the main results from NMA are of great interest as they can limit or enhance the practical implications of the NMA findings. Previous studies have proposed several ways to visualize the results of NMAs,[4] yet no attempt has been made towards integrating the level of confidence in the evidence.

Relevant statistical software is being developed in the last years. The open-source project R is increasingly used

in evidence synthesis, with packages such as meta, rmeta, and netmeta gaining in importance.[5-9] The powerful graphing package ggplot2 greatly facilitates data visualization.[10] The CINeMA software, using the netmeta package and other self-programmed routines facilitates and simplifies the evaluation of the credibility in NMA results.[11] Finally, the shiny package in R makes it possible to build web applications that can take advantage of all R packages and display results interactively.[12] Making good use of these techniques can enhance visualization of the NMA output.

In this article, we present a case where visualization of the NMA results via a Shiny application can provide insight to the evolution of evidence in terms of structure, results, and credibility. We conducted a series of cumulative NMAs at several consecutive time points over the past 40 years. More specifically, we aim to: (a) visualize the evolution of evidence and the level of the confidence regarding the effects of antidepressants using plots which integrate efficacy, acceptability, and confidence in the evidence; (b) examine whether visualization could have facilitated the selection of optimal drugs in earlier years.

2 | METHODS

The protocol for this study has been published.[13] This study used group-level data and did not require approval by an institutional review board. It was registered at UMIN Clinical Trials Registry (identifier: UMIN000031898).

2.1 | Study selection

We used aggregated study-level data identified and collected by GRISELDA (Group of Researchers Investigating Specific Efficacy of individual Drugs for Acute depression).[3] A detailed description of the methods can be found in the protocol.[3,14] Briefly, published and unpublished double-blind RCTs of the acute phase treatment of adult patients (≥18 years old) with a primary diagnosis of major depression were included. The eligible drugs were: agomelatine, amitriptyline, bupropion, citalopram, clomipramine, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, nefazodone, paroxetine, reboxetine, sertraline, trazodone, venlafaxine, vilazodone, and vortioxetine. Only study arms with drugs administered at doses within approved ranges by drug regulatory agencies were included in the analysis.[3]

2.2 | Data extraction

In this study we used two primary outcomes, efficacy (measured as the response rate, that is, the proportion of patients who showed a reduction of at least 50% on validated depression severity scales compared to baseline at a time point as close to 8 weeks as possible between 4 and 12 weeks) and acceptability (measured as the proportion of patients who withdrew early due to any reasons).

Since we aimed to show the evidence evolution, we conducted NMAs at different time periods. Different NMAs included RCTs conducted during different periods, thus we extracted the completion year of each trial. Publication date was used if the trial's completion year could not be identified; and if both were unavailable, the date of approval by drug agencies was used.

The dataset also comprises risk of bias assessments for each included RCT, based on the Cochrane Collaboration risk of bias tool.[15] The risk of bias (low, high, or unclear) was evaluated for 7 domains (sequence generation, allocation concealment, blinding of participants, blinding of therapist, blinding of assessor, selective reporting bias, and attrition bias) for each RCT. The domain-specific judgements were then summarized to obtain a total overall risk of bias for each study. A study was considered at *low risk* if none of the domains were assessed as high risk and three or less were rated as unclear risk; at *moderate risk* if one domain was of high risk, or none was high but four or more were assessed as unclear risk; in all the other situations studies were classified as having *high risk* of bias.[3,14]

2.3 | Data analysis

We programmed a Shiny application that performs NMAs and presents the results interactively. The web app is accessible at https://cinema.ispm.unibe.ch/shinies/GRISELDA/, and several exemplar screen shots from the app are provided in the Appendix. We used the netmeta 1.1-0 package in R (version 3.6.0) to conduct NMA, and to produce the network diagrams, league tables, and forest plots. We used ggplot2 package to draw two-dimensional plots combining both efficacy and acceptability.

In the web app, NMAs can be performed under different conditions. Users can decide whether they want to see the results of an analysis from: (a) both placebo-controlled trials and head-to-head trials, or head-to-head trials only; (b) RCTs completed at any particular time period; and (c) both published and unpublished studies, or published studies only. For each NMA, a random-effects model was used to synthesize odds ratios (ORs)

for both efficacy and acceptability. The transitivity assumption of the entire dataset was previously deemed plausible in the most recent analysis.[3] We examined the comparison-adjusted funnel plot to see if there were significant differences between precise and imprecise trials.

In the *Results* section of this article, we show a series of consecutively conducted NMAs (at 1990, 1995, 2000, 2005, 2010, and 2016, respectively) in order to demonstrate the evolution of evidence, and each analysis includes RCTs completed up to 1 year before that date. These NMAs pertain to the dataset that includes unpublished studies but excludes placebo-controlled trials, as it was also the case in the original publication by Cipriani et al.[3] When interest lies in differences between drugs, placebo-controlled trials can introduce heterogeneity and may compromise the transitivity assumption in the network.[16-20] We also assessed the confidence in the evidence for these NMAs, using the CINeMA (Confidence in Network Meta-Analysis) framework[21-25] through the dedicated software.[11] The confidence in each NMA OR_{AB} between two given drugs A and B was evaluated for six domains: within-trial bias, reporting bias, indirectness, imprecision, heterogeneity, and incoherence. The software required some input in each domain in order to recommend whether there were "major concerns," "some concerns," or "no concerns" for that particular domain. After judgements for all six domains, we summarized the overall confidence in evidence for each NMA OR_{AB} into high, moderate, low and very low. For each drug, we calculated the percentage of the four levels based on all comparisons including that drug, combining both efficacy and acceptability. We therefore produced a level of confidence for the evidence provided for every single drug. More details on CINeMA can be found in References 21-25, and we also provide details of our considerations for assessment in the Appendix. Additionally, as CINeMA assessment requires some threshold values and evaluation rules to be decided, we finalized them through discussions. After determining these rules, the remaining synthesis of confidence in the evidence can be automatically calculated via CINeMA web app, hence one author finally input all the data and got the results.

2.4 | Results presentation and visualization

The Shiny displays the following graphical and numerical information:

1 **Network plots and basic information for both efficacy and acceptability network.** The basic information includes the number of studies, the list of treatments, and some statistics to show the features of the selected network. The sum of within-design Q statistics is used to assess the heterogeneity among direct comparisons (heterogeneity Q statistic), and the between-designs Q statistic is used to assess the inconsistency between direct and indirect evidence (inconsistency Q statistic). The I-squared statistic describes the percentage of variation across studies, that is, due to heterogeneity and inconsistency rather than chance. Since we use random-effects model, the common heterogeneity SD τ is used to indicate the extent of variation among the underlying effects in different studies.

2 **Two-dimensional plots that combines ORs for efficacy and acceptability.** Each node in the plot indicates a drug. The plot has a fixed range on both x and y axis, to accommodate changes over time. Users can click on the plot to see a magnified version based on optimal ranges of x- and y-axis, displayed below the original graph. Placebo is used as the reference if placebo-controlled trials are included, while citalopram as the reference if only head-to-head trials are included. We choose citalopram because it was the most consistently used antidepressant through these years.[26] Moreover, if the input includes placebo-controlled trials and both published and unpublished studies, the plot would have two sets of dots, one for full data and one for published studies only.

3 **Forest plots for both efficacy and acceptability.** The ORs in both efficacy and acceptability between any two drugs are presented using *forest plots* against a common comparator. The reference drug is the same with two-dimensional plots.

4 **League tables**: League tables that can show ORs in both efficacy and acceptability at comparison level are provided for users to download.

5 **Pairwise comparisons for both efficacy and acceptability**: ORs for any two selected drugs as estimated from NMA are displayed on demand.

6 **Comparison-adjusted funnel plots for both efficacy and acceptability**: If placebo-controlled trials are included, the plots show all the RCTs that are directly compared with placebo. If only head-to-head trials are included, the plots show studies directly compared with fluoxetine only, since fluoxetine is the most frequently used active controls over time.

7 **Evidence evolution**: The two-dimensional plots and league tables integrating the efficacy, acceptability and confidence in the evidence for 6 consecutive NMAs at 5-year intervals are presented. For the two-dimensional plots, we use a pie chart as a node for each drug, showing the distribution of the four-level evidence confidence for each drug. For the league tables, we colored each cell in terms of the overall

confidence in evidence between two drugs: green indicated high, blue indicated moderate, yellow indicated low, and red indicated very low confidence in evidence.

In the *Results* section, we present the colored league tables (at comparison level) and the pie charts integrated two-dimensional plots (at drug level) in terms of efficacy, acceptability, and confidence in the evidence, at 5-year intervals since 1990. They reflect the evidence evolution.

3 | RESULTS

3.1 | Included studies and characteristics of the network

The dataset includes 190 head-to-head trials (179 trials with 33 428 patients for efficacy, and 174 trials with 31 596 patients for acceptability, respectively) and 460 placebo-controlled trials; only the results from the former are presented below. The study selection process

(PRISMA flowchart), the references, characteristics and risk of bias assessment for all the included studies are in the Appendix (eFigure 1, eTables 1 and 2). Figure 1 shows the network diagrams for efficacy from 1990 to 2016. The network diagrams for acceptability are shown in the Appendix (eFigure 2).

3.2 | Changes in efficacy, acceptability and confidence in the evidence over time

Figure 2 shows the league tables for efficacy (lower triangle) and acceptability (upper triangle) for each timepoint. In 1990s the evidence for most comparisons was of low or very low confidence, while after 2010 there was moderate or high confidence in the evidence for increasingly more comparisons. eFigures 3 and 4 in the Appendix illustrate the summary ORs for efficacy and acceptability against citalopram at each time point. We also generated the cumulative forest plot for each drug in efficacy compared with citalopram (eFigure 5). For drugs that we could observe for more than 20 years, most of them display a decrease in ORs except fluvoxamine. The

FIGURE 1 Network diagrams for efficacy over the years. The width of the lines is proportionate to the number of studies including that comparison. The size of nodes is proportionate to the number of randomized patients for that arm [Colour figure can be viewed at wileyonlinelibrary.com]

(A)

(B)

(C)

FIGURE 2 Efficacy, acceptability and confidence in the evidence for each comparison over the years. The lower triangle of the table illustrates the efficacy and the upper triangle indicates acceptability. The value in each cell indicate the OR and its 95% CIs for each comparison. It is defined as the treatment listed in the column compared with the drug listed in the row. As a result, for both efficacy and acceptability, ORs≥1 implies that the column-defining drug is favored. Statistically significant results were presented in bold and underlined. The confidence of evidence based on CINeMA for each comparison is shown in different colors in each cell. Green: high confidence; blue: moderate confidence; yellow: low confidence; red: very low confidence [Colour figure can be viewed at wileyonlinelibrary.com]

(D)

(E)

(F)

FIGURE 2 (Continued)

comparison-adjusted funnel plots (eFigures 6 and 7) do not suggest any important association between study results and study precision at any timepoint. The

Appendix details the assessment of the confidence in the evidence (eTables 3–8 for efficacy and eTables 9–14 for acceptability).

FIGURE 3 Two-dimensional plots about efficacy, acceptability and confidence in the evidence from 1990 until 2016. Results are presented as ORs compared with citalopram. Efficacy is shown in x-axis, with ORs≥1 favoring the specific drug, while acceptability is shown in y-axis, with ORs≥1 favoring citalopram. Therefore, the drugs in the right upper corner should be better in both efficacy and acceptability. The node for each drug is shown in terms of pie chart, which indicates the composition of 4-level confidence of evidence of all the comparisons including that drug, of both efficacy and acceptability. Green: high confidence; blue: moderate confidence; yellow: low confidence; red: very low confidence. The drugs in the up-right direction of green dotted lines indicate generally recommendable drugs, while drugs in the down-left direction of red dotted lines indicate the less recommendable drugs at that time point, based on efficacy, acceptability and certainty of evidence. The size of each node is proportionate to the inverse of the width of confidence interval regarding efficacy. The label of a drug name in pink and bold indicates that drug first appears in that network (a relatively new drug). A~F shows the evidences at 6 time points: 1990, 1995, 2000, 2005, 2010, and 2016 [Colour figure can be viewed at wileyonlinelibrary.com]
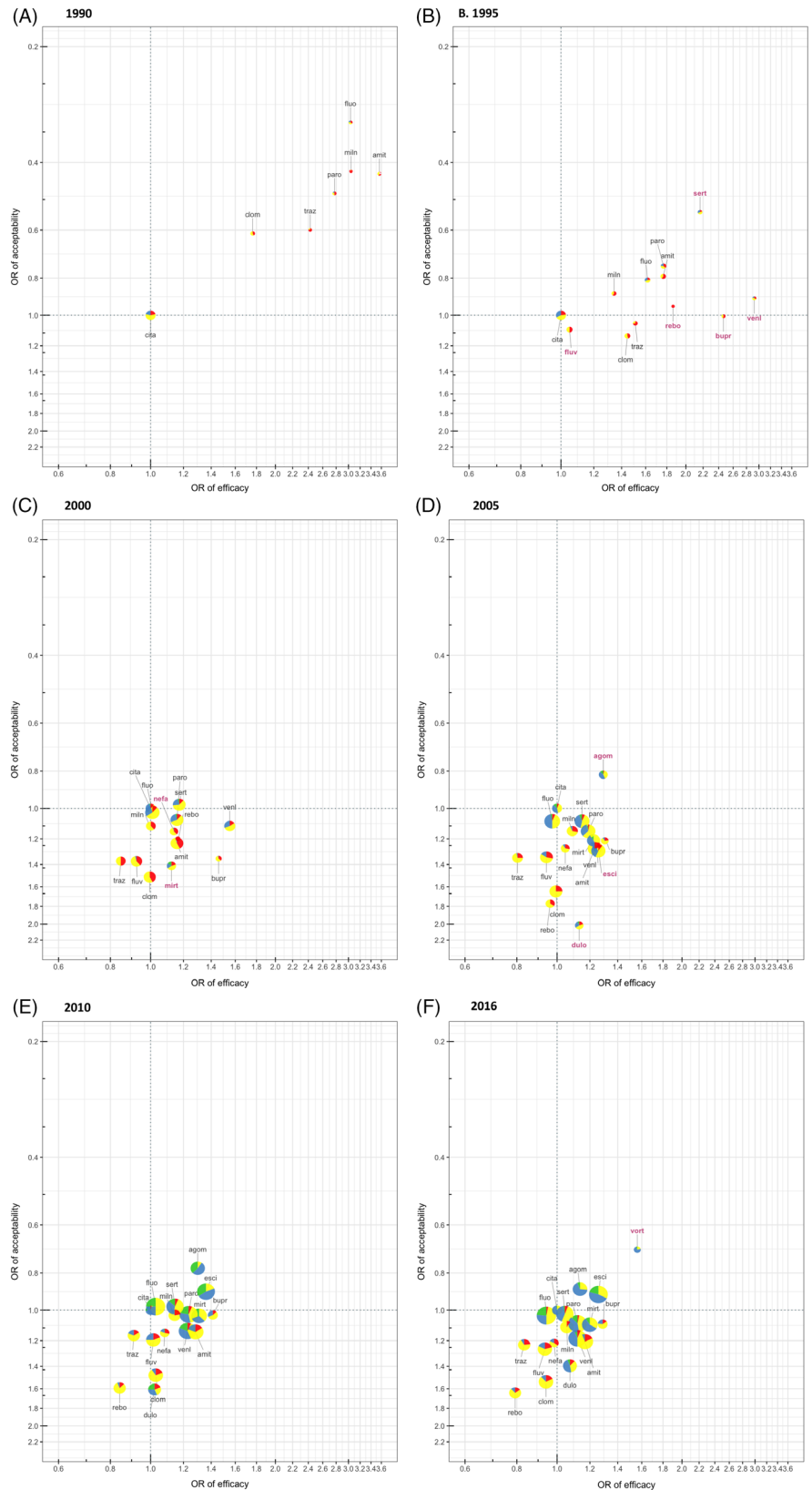
Figure 3 shows the two-dimensional plots at each time point, accommodating effects and the together with the confidence in evidence for each drug in comparison with citalopram. The graphs show that, as expected, precision for ORs improves over time. In 1990, all drugs were apparently more efficacious and acceptable than citalopram, but the confidence in the evidence was in general low due to imprecision as few RCTs were available. In 1995, two new drugs,

venlafaxine and sertraline had relatively higher confidence in evidence compared with other drugs, though half of the evidence still had low credibility and the CIs were very wide. Venlafaxine outperformed in efficacy and sertraline in acceptability. Another two new drugs, bupropion and reboxetine, both had large ORs in efficacy, but the confidence in the evidence was low. In 2000, venlafaxine appeared to outperform all other drugs. Since many drugs in 2000 had lower ORs in efficacy compared to 1995, their effects are closer to citalopram's and hence difficult to visually distinguish between them. At the same time, trazodone, fluvoxamine and clomipramine had worse efficacy, tolerability and credibility in the evidence compared to other drugs. In 2005, new drugs were launched into the market, among which agomelatine stood out for its performance in efficacy and acceptability with moderate confidence in the evidence. Several drugs performed equally well, such as escitalopram, mirtazapine and paroxetine. Trazodone, fluvoxamine, clomipramine, and reboxetine were found to have low efficacy, acceptability with low confidence in the evidence. In 2010, with further accumulation of studies, the confidence in the evidence improved overall. Agomelatine and escitalopram were presented as good options for both efficacy and acceptability. Finally, in 2016, vortioxetine was introduced and appeared to have the best efficacy and tolerability, however, the precision was very low. Our results implied that the findings for vortioxetine might be too optimistic and we should wait some years to see its real effect.

Other than the figures displayed in this article, users can also explore the analyses which involve placebo-controlled trials in the web app. The comparison-adjusted funnel plots including placebo-controlled trials for the whole dataset indicate statistically significant asymmetry (P-value for Egger's test is .0025 in both efficacy and acceptability). However, since the plots are not extremely asymmetrical, the small P-value may due to the large number of RCTs. If changing the analysis to pooling more recent studies (eg, after 1990), the test of asymmetry becomes insignificant. It suggests that the small-study effect may exist in old RCTs. The two-dimensional plots imply that the differences between active drugs are smaller at any time point when placebo-controlled trials are included. This phenomenon has been extensively discussed and possible explanations were given in a previous study.[20] Users can also get an appreciation of the magnitude of a potential publication bias when placebo-controlled trials are included from the two-dimensional figures.

4 | DISCUSSION

To our knowledge, this is the first study to investigate and visualize the evolution of evidence concerning antidepressants in treating patients with acute phase major depression in the past 40 years, based on efficacy, acceptability, and confidence in evidence using cumulative NMAs. We developed a visualization tool via Shiny in R to facilitate the understanding of a massive amount of information produced about multiple analyses: multiple drugs, types of studies and time points for two outcomes and account of the credibility of the results. The web application captured the dynamic changes in efficacy and acceptability over time as new drugs were launched into the market and more and more trials completed. We observed that, initial results usually exaggerated the performance of newly approved drugs, and it took time to see the real effects, as was the case of sertraline, venlafaxine, bupropion, and many others. As the network became better populated with more studies and better connected over time, the overall confidence in the evidence improved.

This is the first study to display the confidence in evidence together with the estimates of efficacy and acceptability for NMAs. Confidence in evidence is just as crucial to clinical decision making as effect size estimates themselves. In order to integrate three levels of information concurrently, we modified the league tables and two-dimensional figures. We used different colors of the cell to show the confidence of evidence for each comparison. If there are two or three candidate drugs to select from for a particular patient, it would be easy to check the relative effect and certainty of evidence between them in the league table. However, it is not easy to see the general performance for each drug from the league table, since it is comparison-based. Several studies adopted a two-dimensional graph to combine the efficacy and acceptability ORs for each drug.[3,27] In addition to these conventional two-dimensional graphs, we presented the certainty of evidence for each drug by illustrating the composition of four-level confidence in a small pie chart. It allows us to easily see that even though some drugs showed good efficacy and acceptability, if the evidence appeared to have low certainty, the recommendation of those drugs should be made with caution.

As NMA is a method to synthesize massive data but yet produce a large amount of results, it is imperative to find a way to display their results more efficiently for physicians to comprehend. A powerful R package Shiny makes it easy to build interactive web apps straight from R. Within the Shiny application that we have developed, users can change some parameters and view the results of their own interest and to answer their own clinical questions.

We observed that among drugs whose evolution of evidence we could observe for more than 20 years, almost all of them (except fluvoxamine) showed noticeable

reduction in the estimate of efficacy after being launched into the market, and it usually took more than 10 years to become stable. It indicates that new drugs tend to show overly enthusiastic performance when they are launched. This phenomenon has been observed in many other studies involving different medical specialties.[28-30] It was named variously, such as "wish bias,"[29] "the fading of reported effectiveness,"[30] or "novel agent effects."[31] The inflated initial effects for new drugs were considered to be related to highly selected patients, small sample sizes, selective reporting, and publication bias in the early stage trials.[29,32,33] As a result, we should be particularly careful with the new drugs. Strategies such as improving the design, prespecifying rigorous analysis plans in the protocols, or encouraging prompt replications with bigger sample sizes, should be promoted for future early trials of new drugs.[32,34] Meanwhile, accelerating evidence accumulation through evidence synthetic methods such as NMAs may also help.

Rouse et al recently conducted similarly conceived retrospective NMAs for medications for open-angle glaucoma at four to five-year intervals since 1991, in which the most efficacious drug remained to be prostaglandins since 1999.[35] In our analysis by contrast, there were more dramatic changes in the recommendable drugs. Our results also show that the evidence can easily get out of date. It therefore follows that it is highly necessary to update evidence synthesis in a timely manner. Conducting NMAs from scratch is time-consuming.[36,37] However, if we conduct a prospective NMA according to the pre-specified procedures, we only need to update the dataset regularly. NMA may therefore be the most optimal method to synthesize evidence dynamically. In fact, the concept of living NMA, or sequential NMA, methodologically similar with retrospective cumulative NMA but with due attention to multiple testing issues, has already been proposed.[1,38-40] It can provide the theoretical basis for living systematic reviews and living guidelines for future practices of evidence-based medicine.[37,41-43]

4.1 | Limitations

Our study has several limitations. First, the confidence in evidence assessment via CINeMA inevitably involves subjective decisions, such as how to summarize average risk of bias or how to set the margin of equivalent effects. It is possible that different decisions might lead to different appraisal of the confidence in the evidence. To minimize the risk of post-hoc findings, we prespecified all the conventions in our protocol and we tried to make them clinically reasonable through discussions. Second, abstracting and summarizing the characteristics of the

studies and lumping them across comparisons and drugs comes at the price of losing some important information. For example, aggregation of the level of confidence to present drug-level judgements can conceal which comparisons and which confidence domains are responsible for the low confidence. Third, the acceptability of drugs was based on the overall discontinuation rate from RCTs, which might be slightly different from the concept of tolerability. Besides, RCTs are not the best method to evaluate long-term tolerability or adverse effects. Fourth, we conducted NMAs every 5 years, which was somewhat arbitrary. However, we supposed that an interval of 5 years may suffice to capture the change in the evidence. On the web a newly developed interactive app enables analyses at any time points, though confidence assessment was only at 5-year interval. Finally, although our study illustrates the efficacy, acceptability, and confidence in the evidence simultaneously, the real recommendations to instruct clinical practice should also take cost-effectiveness analysis or and specific side effects into consideration.

5 | CONCLUSION

Our interactive visualization via Shiny in R captures the dramatic changes in evidence about the effects of antidepressants in the last 40 years, showing that initial results of new drugs are often exaggerated and that the overall confidence in the evidence improves over time. As a result, we should stay conservative towards new drugs, but speed up evidence accumulation and always keep the synthesized evidence up to date.

CONFLICT OF INTEREST

CONTRIBUTORS

Yan Luo and Toshi A. Furukawa designed the study. Yusuke Ogawa collected data, Anna Chaimani and Yan Luo conducted statistical analyses. Yan Luo constructed the web application. Toshi A. Furukawa and Georgia Salanti gave suggestions for analytical plans. All the authors participated in interpretation of the results. Yan Luo drafted the manuscript and all authors critically revised the manuscript and approved the final version. All authors gave final approval of the version to be published.

ETHICS STATEMENT

This study does not require institutional review board approval. We have registered it in the University hospital Medical Information Network Clinical Trials Registry (UMIN-CTR): UMIN000031898.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in https://github.com/y-luo06/cNMA_of_antidepressant.

ORCID

Yan Luo https://orcid.org/0000-0002-5271-5126

REFERENCES

1. Nikolakopoulou A, Mavridis D, Furukawa TA, et al. Living network meta-analysis compared with pairwise meta-analysis in comparative effectiveness research: empirical study. *BMJ*. 2018;360:k585.
2. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet*. 2009;373(9665):746-758.
3. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 2018;391 (10128):1357-1366.
4. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS One*. 2013;8(10):e76654.
5. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/. Published 2019.
6. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. 2019;22:153-160. https://doi.org/10.1136/ebmental-2019-300117.
7. Lumley T. rmeta: Meta-Analysis. R package version 3.0. Retrieved from https://CRAN.R-project.org/package=rmeta. Published 2018.
8. Rücker G, Krahn U, König J, Efthimiou O, Schwarzer G. netmeta: Network Meta-Analysis using Frequentist Methods. R package version 1.1–0. Retrieved from https://CRAN.R-project.org/package=netmeta. Published 2019.
9. Schwarzer G, Carpenter JR, Rücker G. *Meta-Analysis with R*. Switzerland: Springer International Publishing; 2015.
10. Wickham H, ed. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag; 2016.
11. CINeMA. CINeMA: Confidence in Network Meta-Analysis [Software]. Institute of Social and Preventive Medicine. Retrieved on August 2019 from cinema.ispm.unibe.ch. Published 2017.
12. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. shiny: Web Application Framework for R. R package version 1.4.0. Retrieved from https://CRAN.R-project.org/package=shiny. Published 2019.
13. Luo Y, Chaimani A, Kataoka Y, et al. Evidence synthesis, practice guidelines and real-world prescriptions of new generation antidepressants in the treatment of depression: a protocol for cumulative network meta-analyses and meta-epidemiological study. *BMJ Open*. 2018;8(12):e023222.
14. Furukawa TA, Salanti G, Atkinson LZ, et al. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ Open*. 2016;6(7):e010919.
15. Higgins JPT, Green Se. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. Version 5.1.0 [updated March 2011] ed: The Cochrane Collaboration; 2011.
16. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome?. The effects of placebo control and treatment duration in antidepressant trials. *Psychother Psychosom*. 2009;78(3):172-181.
17. Sinyor M, Levitt AJ, Cheung AH, et al. Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. *J Clin Psychiatry*. 2010;71(3):270-279.
18. Furukawa TA, Cipriani A, Atkinson LZ, et al. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. *Lancet Psychiatry*. 2016;3(11):1059-1066.
19. Furukawa TA, Cipriani A, Leucht S, et al. Is placebo response in antidepressant trials rising or not? A reanalysis of datasets to conclude this long-lasting controversy. *Evid Based Ment Health*. 2018;21(1):1-3.
20. Salanti G, Chaimani A, Furukawa TA, et al. Impact of placebo arms on outcomes in antidepressant trials: systematic review and meta-regression analysis. *Int J Epidemiol*. 2018;47(5):1454-1464.

21. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, et al. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS Med*. 2020;17(4):e1003082.

22. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PLoS One*. 2014;9(7):e99682.

23. Papakonstantinou T, Nikolakopoulou A, Higgins JPT, Egger M, Salanti G. CINeMA: Software for semi-automated assessment of the Confidence In the results of Network Meta-Analysis. Retrieved on 14 November 2019 from https://cinema.ispm.unibe.ch/model/CINeMA_manual.pdf

24. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, et al. Assessing Confidence in the Results of Network Meta-Analysis (Cinema). *bioRxiv*. 2019:597047.

25. Papakonstantinou T, Nikolakopoulou A, Higgins JPT, Egger M, Salanti G. CINeMA: software for semiautomated assessment of the confidence in the results of network meta-analysis. *Campbell Syst Rev*. 2020;16:e1080.

26. Luo Y, Kataoka Y, Ostinelli EG, Cipriani A, Furukawa TA. National Prescription Patterns of antidepressants in the treatment of adults with major depression in the US between 1996 and 2015: a population representative survey based analysis. *Front Psych*. 2020;11:35.

27. Mavridis D, Giannatsi M, Cipriani A, Salanti G. A primer on network meta-analysis with emphasis on mental health. *Evid Based Ment Health*. 2015;18(2):40-46.

28. Trikalinos TA, Churchill R, Ferri M, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol*. 2004;57(11):1124-1130.

29. Barbui C, Cipriani A, Brambilla P, Hotopf M. "Wish bias" in antidepressant drug trials? *J Clin Psychopharmacol*. 2004;24(2):126-130.

30. Gehr BT, Weiss C, Porzsolt F. The fading of reported effectiveness. A meta-analysis of randomised controlled trials. *BMC Med Res Methodol*. 2006;6:25.

31. Salanti G, Dias S, Welton NJ, et al. Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med*. 2010;29(23):2369-2383.

32. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology*. 2008;19(5):640-648.

33. Tajika A, Ogawa Y, Takeshima N, Hayasaka Y, Furukawa TA. Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. *Br J Psychiatry*. 2015;207(4):357-362.

34. Ioannidis JP. Research accomplishments that are too good to be true: reply to ting. *Intensive Care Med*. 2014;40(3):468.

35. Rouse B, Cipriani A, Shi Q, Coleman AL, Dickersin K, Li T. Network meta-analysis for clinical practice guidelines: a case study on first-line medical therapies for primary open-angle glaucoma. *Ann Intern Med*. 2016;164(10):674-682.

36. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med*. 2007;147(4):224-233.

37. Elliott JH, Turner T, Clavisi O, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med*. 2014;11(2):e1001603.

38. Nikolakopoulou A, Mavridis D, Egger M, Salanti G. Continuously updated network meta-analysis and statistical monitoring for timely decision-making. *Stat Methods Med Res*. 2018;27(5):1312-1330.

39. Wetterslev J, Jakobsen JC, Gluud C. Trial sequential analysis in systematic reviews with meta-analysis. *BMC Med Res Methodol*. 2017;17(1):39.

40. Vandvik PO, Brignardello-Petersen R, Guyatt GH. Living cumulative network meta-analysis to reduce waste in research: a paradigmatic shift for systematic reviews? *BMC Med*. 2016;14:59.

41. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schunemann HJ. Living systematic review N. living systematic reviews: 4. Living guideline recommendations. *J Clin Epidemiol*. 2017;91:47-53.

42. Elliott JH, Synnot A, Turner T, et al. Living systematic review: 1. Introduction-the why, what, when, and how. *J Clin Epidemiol*. 2017;91:23-30.

43. Crequit P, Martin-Montoya T, Attiche N, Trinquart L, Vivot A, Ravaud P. Living network meta-analysis was feasible when considering the pace of evidence generation. *J Clin Epidemiol*. 2019;108:10-16.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.