# Actively Living Network Meta-Analysis

*Georgia Salanti and Adriani Nikolakopoulou*

Table of Contents

## 1 *Summary of the framework*

**What is already known:**

- Network meta-analyses are increasingly becoming the norm in evidence synthesis for comparative effectiveness research and living systematic reviews have been proposed as a mean to providing timely evidence.

**What is new:**

- 'Actively Living Network Meta-Analysis' combines ideas of living systematic reviews and network meta-analysis with an intention to inspire the production of further studies and inform their designs according to existing evidence gaps

- Methods underpinning 'Actively Living Network Meta-Analysis' differ from those in conventional network meta-analysis in two major aspects; a) inference about treatment effects is drawn considering the frequent updating of the evidence b) the need for further research and updating are expressed quantitatively via sample-size calculations for future trials.

- Input from the end-users such as guideline developers is indispensable when planning and undertaking an 'Actively Living Network Meta-Analysis'

-

**What are the implications:**

- 'Actively Living Network Meta-Analysis' has numerous advantages compared to all other evidence synthesis forms: increased precisions, timeliness and speeding-up the production of high-quality conclusions about the effects of interventions. Downsides include the greater workload, complex methodology, the need for wider cross-fields collaboration

- Planning future studies using 'Actively Living Network Meta-Analysis' results in a considerably lower sample size to reach conclusions.

- The major constraint in adopting 'Actively Living Network Meta-Analysis' to plan future studies is the current paradigm of funders of clinical research and stakeholders as it represents a methodological frontier

## 2   Introduction and rational

Considerable progress in computing tools and statistical methodology in the last decade have transformed the landscape in evidence synthesis and enabled systematic reviewers to address complex contemporary health policy questions. Evidence synthesis tools and frameworks like living systematic reviews and network meta-analysis are among the recent ground-breaking changes in the field.

Living systematic reviews (LSR) have been introduced two years ago [1]. They are systematic reviews that are continuously updated with new data as they become available. Their uptaking can be supported by machine-learning techniques for identifying, screening and extracting data from studies, by data-sharing platforms and online technologies that facilitate collaboration. A few examples have already appear in the literature [2-4], and the Cochrane Collaboration is currently piloting living systematic reviews for some key research questions of priority.

Network meta-analyses (NMA) and indirect comparison appear frequently now in the medical literature (Petropoulou et al) and their role in guideline development [5], health technology assessment [6] and even in regulatory setting [7] is increasingly important. The first attempt to take NMA into a living setting has recently appear [8] and was received with enthusiasm [9]. While the statistical methodology underpinning NMA has been presented in many tutorials [10], its 'living' version represents a methodological frontier and requires extra considerations that haven't been discussed in the literature.

The key feature of a living systematic review is the frequent update of the evidence with new studies. This feature could be characterized as "passive"; reviewers simply wait for new data to become available for inclusion and then they decide whether further updates are needed using some decisions rules, for example as described in [11,12]. This opportunistic updating could be improved by making specific recommendations about what sort of research is needed to improve the output of a LSR  by identifying gaps in the evidence and could help funders of clinical research to minimizing research waste [13].

The role of evidence synthesis in directing future research has been prominent in opinion articles [14] but empirical studies have shown that new randomized controlled trials (RCTs) do not consider, in their majority, existing meta-analyses. Out of 24 trialists whose studies were included in updates of Cochrane reviews, about half reported that they were aware of the relevant reviews when they designed their trial and only one in three designed their study taking into account the review [15].

When it comes to using the data of meta-analysis to inform sample size calculations, empirical evidence is even more condemning. Only one in five RCTs funded by NIHR HTA between 2006 and 2008 used the meta-analysis to calculate the sample size in the new trial [16]. In a review of 446 trial protocols submitted to the UK ethics committees in 2009, less than 1% reported that they used a meta-analysis to set the treatment difference sought [17]. One in ten trials published in leading medical journals in 2007 used meta-analysis to inform their design but only one in 25 use of meta-analysis in a sample size calculations [18].

In this article we introduce the concept of 'Actively Living Network Meta-Analysis'(AL-NMA); a living network meta-analysis that actively makes specific suggestions about the need of further studies to answer the research question they address. AL-NMA aims to provide up-to-date evidence while optimizing research investment by detecting evidence gaps and indicates an optimal new study that will update the existing evidence to best serve public health decision making. This new evidence synthesis tool best supports ideas of 'conditional licensing' and 'conditional reimbursement' where regulatory agencies and HTA bodies make decisions conditional on additional evidence made available in the near future. It is also in line with international consensus about the need to reduce waste and increase value in research [13].

In this article we describe the main methodological characteristics of AL-NMA placing emphasis on those quantitative characteristics that are different to a conventional NMA or a living systematic review as previously described [1,19]. We present two examples from existing network meta-analyses and we provide results from a survey that aims to explore the acceptability of the paradigm shift that an AL-NMA entails. We confine ourselves to the case where the motivating research question relates to post-marketing decisions, such as guideline development and reimbursement. Because cost-effectiveness analysis as part of the reimbursement decision is specific to the time and setting of the decision-making process, we consider information relevant to treatment effectiveness, safety and tolerability, rather than costs of treatments.

## 3    *Methods for Actively Living Network Meta-Analysis*

The lifecycle of an AL-NMA is presented in Figure 1. The main methodological considerations underpining the decisions involved in the flowchart are included in Box 1

**Figure 1 The main steps in Actively Living Network Meta-Analysis. Decisions in each step should use pre-defined criteria and processes as described in Box 1**

4
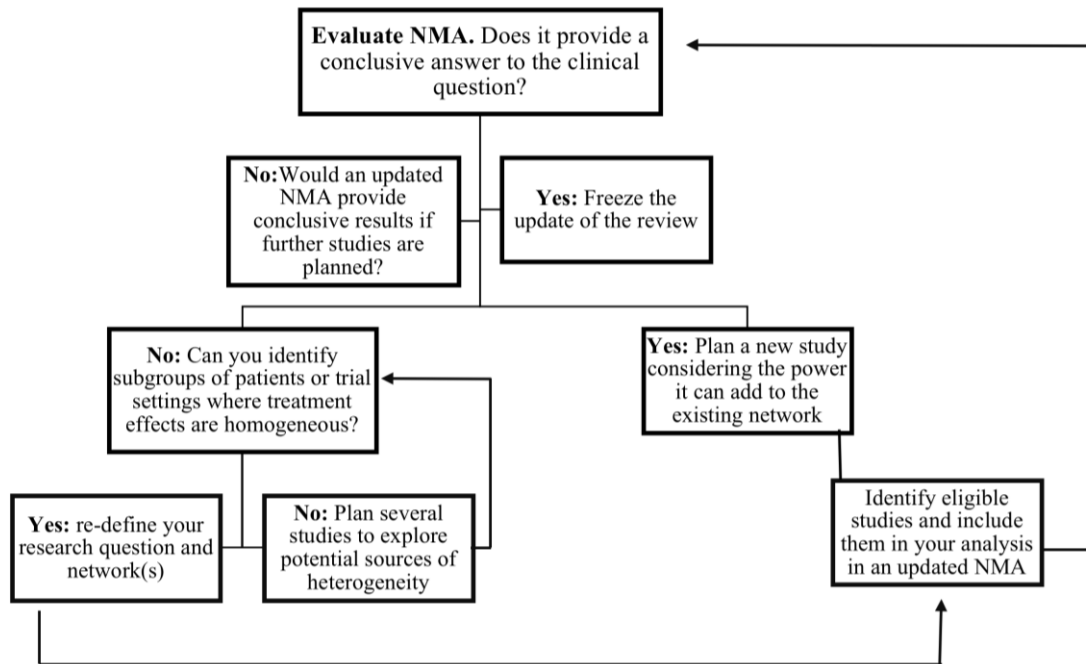
The starting point is a NMA about the relative efficacy and safety of competing interventions for the same health condition. We assume that the assumptions underlying NMA have been deemed reasonable; heterogeneity is not very large and the assumption of transitivity is defendable. There are no hard lines as to what consists too high heterogeneity and when the transitivity assumption might hold. Resorting into tests has been shown to be misleading and the power of tests for heterogeneity and inconsistency has been shown to be low (reference). Clinical and epidemiological judgement should be combined with the application of several statistical tools to decide whether the starting NMA complies with the methodological standards as described in several tutorials [19].

The first decision that needs to be taken is whether the existing evidence answers the starting research question in a conclusive manner. If not, it needs to be decided whether accumulation of additional data is likely to provide a conclusive answer, and if yes, what is the optimal design for a new trial. If further studies are unlikely to lead into conclusiveness, then exploration of the heterogeneity of treatment effects (by potentially planning further experiment) could be a way forward. Finally, new data from the planned study and other relevant studies are embedded in NMA into an updated version. The methods used in each step need to be detailed in the protocol of the AL-NMA and are summarized in Box 1. We assume that the whole process is undertaken by a collaborative panel that includes clinical and methodological experts, information specialists and patient representatives.

1. **Describe the criteria that would be employed to characterize the evidence provided by NMA as conclusive.** Important considerations include a) choosing treatment comparisons of interest (termed "targeted comparisons") b) weighting positive and negative outcomes c) defining worthwhile relative treatment effects against which the precision of the summary effects would be judged d) evaluate the risk of bias in the body of evidence e) evaluate evidence synthesis assumptions (homogeneity and consistency).

2. **Describe how heterogeneity in the targeted comparisons will be measured and accounted for.** The assumptions about the heterogeneity (e.g. whether a single heterogeneity parameter is employed for all treatment comparisons or not, the use of informative priors at early phases of the network etc.) have important impact on the estimated or assumed heterogeneity and on the decisions about the relevance of planning further studies.

3. **Describe realistic options for trial scenarios.** The feasibility, cost and ethics underpinning all treatment options shall be considered when setting up those scenarios.

4. **Describe the statistical methods employed to synthesize data.** Besides standard considerations in NMA, describe whether you will account for repetitive (sequential) analyses of the cumulative evidence and if yes how.

5. **Try to foresee which items of the methodology are likely to change over time** and how you will deal with such changes to limit post-hoc decisions. Changes might be triggered by the marketing of new relevant interventions or withdrawal of some of them, expiring patents that shift considerations bout the cost of drugs, changes in heterogeneity because of time-dependent effect modifiers etc.

**Box 1 Methodological aspects to be included in the protocol of an Actively Living Network Meta-analysis**

*3.1    Is the current evidence conclusive?*

In our context, we define as conclusive the evidence that can be used to make confident decisions for reimbursement or for clinical practice. The protocol of an AL-NMA should include the outcomes and treatment comparisons that are most important for decision making and for which conclusiveness is to be judged. NMA typically involves legacy treatments that provide indirect evidence but are not of interest (e.g. placebo). In the case of reimbursement, the comparisons of interest (termed here "targeted comparisons") might be easy to define, but when it comes to guideline development, a panel of experts and patient representatives needs to define what has been

described as 'the decision set' of treatments [20]. Depending on the context this is likely to be a long list of treatment comparisons or even the entire network. In AL-NMA new interventions might become available with time and hence adaptations about the targeted comparisons of interest would be needed.

Then the GRADE system can be used to grade the evidence. Two approaches have been proposed so far to evaluate the quality of the evidence from NMA but they haven't been tested or compared in practice [21,22]. They both address the standard GRADE items; bias (risk of bias in the included studies and publication bias), indirectness, imprecision and heterogeneity but also make considerations about the plausibility of the assumptions of NMA.

An alternative to GRADE for inferring about conclusiveness is to associated the evidence with the confidence in the decision making as estimated in complex decision-analytic models. Multiple outcomes are typically considered in systematic reviews and their trade-off is taken into account when deciding about the strength of recommendations. Methods to summarize data across multiple treatments and multiple outcomes can be very informal (such as those employed in the GRADE system) or involve a full decision analysis [23].

An important component in the process of characterizing the evidence as conclusive is the uncertainty around the meta-analytical relative treatment effect. Sheer statistical significance against the null hypothesis is not always of interest and the clinical importance of the findings is increasingly considered when interpreting the results from a meta-analysis. Methods that combine the clinical importance of the findings with benefit-harm trade-offs, such as the 'worthwhile effects' are most relevant in guiding decisions about conclusiveness of the evidence [24,25]. In these approaches, quantitative considerations about the conclusiveness of the evidence do not judge whether relative treatment effects are different from the null, but use the 'worthwhile effects' as benchmark; the minimum effect size on the primary beneficial outcome that outweighs risk, costs and inconvenience. A summary treatment effect for specific comparison can be characterized as imprecise if the confidence intervals is too wide and includes the worthwhile effect. In NMA where several comparisons are of interest, evidence can be collectively characterized as imprecise is joint regions of effects or the ranking of the treatments is imprecise.

*3.2    Is conclusiveness likely to be achieved if additional studies are planned?*

Frequent reasons for inconclusive evidence are the high risk of bias in the existing NMA, imprecision in the summary effects (i.e. few, small studies) and variability in the true relative

treatment effect (heterogeneity). Studies can be planned to resolve or attenuate these shortcomings when added to the existing body of evidence. For example, if the high risk of bias in some of the included studies is the major concern in the existing NMA, planning a new adequately powered study with low risk of bias to be added to the network or substitute studies at high risk of bias will improve the overall credibility of the evidence and might lead to conclusive evidence.

Although planning more studies seems to be the natural reaction to inconclusive evidence, it is not always the optimal way forward. If the reason for inconclusiveness is publication bias, then planning a new study would not improve confidence in the existing results. A frequent cause of inconclusiveness for evidence synthesis is imprecision in the summary treatment effects. In that case the added power of a further study highly depends on the variability of the true relative treatment effect. A key statistical concept in planning future studies is the conditional power; that is the power of the meta-analysis when a new study is added rather than the power of a new study in isolation. If heterogeneity is high, the uncertainty about the true effect in a new study is large and hence the contribution of a single study to the conditional power, even of a very large one, is expected to be very low [26–28].

There are several hurdles when deciding whether the present heterogeneity renders the undertaking of future studies irrelevant. Estimation of heterogeneity requires several studies to be available, and this might not be the case in the early phases of the AL-NMA. Heterogeneity cannot be estimated for comparisons for which no direct evidence is available. In such cases, a common heterogeneity parameter for all comparisons in the network is assumed, but might not represent the anticipated heterogeneity in the comparison of interest. Empirical studies have provided estimates of the heterogeneity variance specific to the outcome and treatment comparison and can be used to inform heterogeneity when data is scarce.

Although planning several smaller studies instead of a single larger one could improve conditional power in the presence of important heterogeneity, it is advisable to make all efforts to understand its cause before proceeding. Network meta-regressions and subgroup analyses can be part of exploratory analyses. Identification of patient subgroups within which the targeted treatment effects are more homogeneous could help narrowing down the research question and proceed with AL-NMA in specific patient populations. However, this would typically require individual patient data at the evidence synthesis level. If heterogeneity is large and remains unexplained even after efforts to pinpoint sources of variability, then future studies can be planned with an aim to provide

information about the effect of treatment in various populations and settings and hence reveal the source of heterogeneity.

### 3.3   How to plan a future study in order to render the existing evidence conclusive

When a network meta-analysis is deemed inconclusive and the planning of future studies is desirable, specific recommendations about what sort of studies should be planned next need to be made. Studies can be planned to lower the risk of bias for particular comparisons, to explain heterogeneity or to inform outcomes and comparisons for which evidence is imprecise.

Within the AL-NMA framework, studies are not considered as stand-alone experiments but they are seen as group-sequential additions to the existing evidence. The power and findings of each study are not of interest; the conditional power of the network meta-analysis when the study is added and the summary effect are of importance. Consequently, when the network meta-analysis is deemed inconclusive because of imprecision, sample size calculations should be based on the conditional power or improved precision of an updated meta-analysis. Two related methods have been presented about sample size calculations for a new study based on NMA and extend the work initially done for pairwise meta-analysis meta-analysis [26,27]. A key issue of the methodology is the realization that adding data to a part of the network impacts on all comparisons. Consequently, we can improve precision on a targeted comparison, say treatment A versus B by adding data on another comparison, e.g. A versus C.

The process starts by setting an 'optimal' precision for the targeted comparisons and candidate study scenarios according to pre-defined criteria discussed in section 3.1 and **Error! Reference source not found.** (e.g. so that all worthwhile effects are marginally excluded from the confidence intervals). This optimal precision can be achieved by studies that have various designs (number of arms and interventions involved). The design that requires the least sample size to achieve the optimal precision is that of a multi-arm study involving all treatments in the comparison of interest. As this might not be a realistic option, a set of trial scenarios need to be defined considering practical, ethical and economic aspects of each design. For example, the standard treatment for a condition might not be part of the targeted comparisons, but feasibility, cost and recruitment rates might be better than those for a new and expensive treatment. Sample size calculations can then be carried out for all trial scenarios using the methodology described in [28,29]. It is expected that trial scenarios that include treatments located away from the targeted comparisons would require larger sample size compared to comparisons that are closer to achieve the same level of precision.

9

Additionally, the required sample size is larger for trial scenarios for which larger heterogeneity is expected. The estimated sample size per trial scenario should then be combined with the feasibility, acceptability and cost of each scenario to decide which trial should be done next. Sample size estimation can be more formally combined with the cost in an expected value of sample information framework [30,31]. However, such models require input that is highly specific to the country for which decisions are to be made (i.e. utility measures and costs).

We envision that recommendation for optimal study design to serve decision making as described above would ultimately be adopted by researchers and users of evidence syntheses, but it is the funders of clinical trials who will determine whether such evidence-based planning of clinical trials would become popular in the future. If AL-NMA is undertaken by a large consortium, it is possible that this part of the circular process has high chances of success.

## 3.4   Regular updating of the evidence and re-evaluation of conclusiveness

Once new evidence becomes available, this shall be incorporated in an updated NMA. This new evidence would comprise data from relevant studies that might have become available and possibly data from the optimally designed study, had this been realized. Several tools can be employed to speed-up and decrease the workload in the process of searching and screening for or extracting data from relevant studies (references).  At the end of each update, the process returns to the evaluation of the new evidence against the conclusiveness criteria (Figure 1).

Regular update of the NMAs produces several estimates of the relative treatment effects. Inference about imprecision against the worthwhile effects (or any other pre-defined benchmark values) needs to be repeated after each update. There has been a long debate as to whether adjustment for multiple testing is needed in a regularly updated meta-analysis. Opponents of the adjustment suggest that evidence synthesis aims at estimating effects rather than testing. However, the precision around the estimates is of importance when judging conclusiveness. In contrast to conventional NMA, AL-NMA aims to make decisions about whether further studies are needed and if yes how they should be designed. In this context, we believe that type I error rate needs to be controlled for.

Sequential methods for pairwise meta-analysis have been developed in this spirit [32,33]  and have been extended to NMA context [34]. The method described by Nikolakopoulou can be used to construct stopping boundaries (for efficacy or futility) for the targeted comparisons; or equivalently to construct 'repeated' confidence intervals that account for the sequential nature in the accumulation

of the data. A few other methods have been proposed for pairwise meta-analysis (name them and references) but haven't been extended to the case of NMA.

While NMA is expected to improve precision compared to pairwise meta-analysis, sequential methods for inference in NMA as part of AL-NMA will decrease precision particularly at the early phases of the network putting a hurdle in declaring conclusiveness too early.

## 4    Illustrative application of Actively Living Network Meta-Analysis

### 4.1    Are second generation antipsychotics more effective than Haloperidol?

There is an ongoing debate whether newer and much more expensive second-generation antipsychotics are better than first-generation antipsychotics. The first-generation compound haloperidol had been the standard antipsychotic for decades and olanzapine is among the most frequently prescribed second-generation antipsychotics.  A network meta-analysis in 2013 [35] indicated Amisulpride, Clozapine, Olanzapine and Risperidone as having the best efficacy-acceptability profile and being significantly more efficacious than Haloperidole.

We assume that this network has being planned as living NMA and inference would have been made using sequential monitoring and conclusions about relative treatment effects. An informative prior is being used for heterogeneity according to empirical evidence. Figure 2 presents the boundaries for pairwise and network meta-analyses for the comparisons against Haloperidole. In agreement with the original article, the higher efficacy of Amisulpride, Clozapine and Olanzapine over Haloperidol would have been established even after accounting for multiple testing in an NMA (Figure 2, panels, a,b,d). However, contradicting Leucht et al, the comparison between Risperidone and Haloperidole is not conclusive after adjusting for repeated updating (Figure 2 panel c).  Leucht and colleagues also claimed higher efficacy for clozapine compared to other efficacious second-generation antipsychotics such as amisulpride. If this NMA had been designed as a living NMA, this conclusion would have been challenged. The sequential monitoring shown in Figure 6 suggests that the evidence regarding the comparison of clozapine versus amisulpride is inconclusive regarding efficacy.

Figure 2 exemplifies the advantage of living NM over living pairwise meta-analysis; only evidence about Clozapine would have been conclusive if pairwise meta-analysis had been considered. A simple pairwise meta-analysis of the eleven RCTs directly comparing Olanzapine to

Haloperidole would not have been enough to establish a conclusion whereas including indirect evidence from 167 studies included in the network suggests a significant finding.

**Figure 2 Sequential monitoring for cumulative head-to-head and network meta-analysis between antipsychotics. Heterogeneity standard deviation is informed by respective empirical distributions throughout the exercise. Heterogeneity has been assumed constant and set equal to the median predicted value as estimated from empirical evidence.**
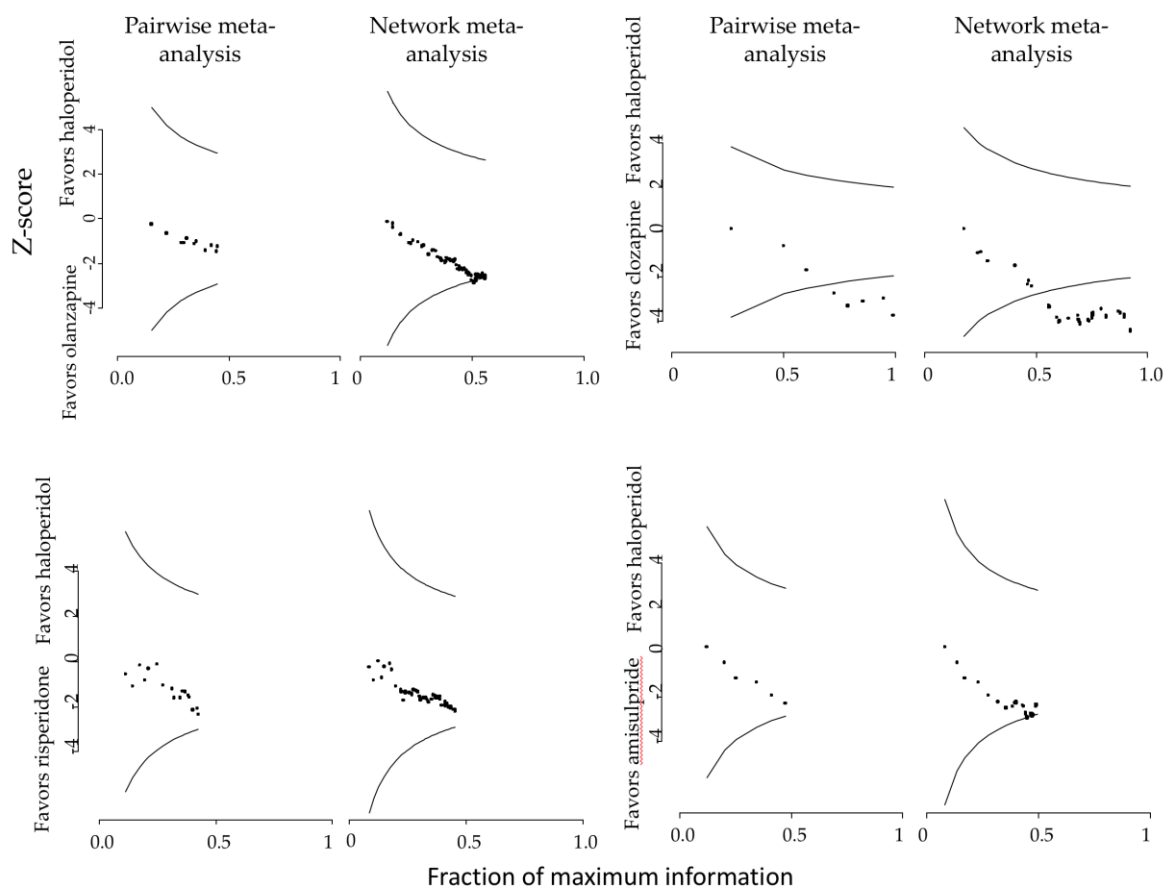


## 4.2 Conventional and biologic DMARDs for rheumatoid arthritis: planning the next trial

Comparison of biologics and conventional DMARDs has been the subject of several NMAs. The lack of direct evidence between new biologic DMARDs renders the use of indirect comparison imperative. The most recent publication suggested that in methotrexate-naïve patients, a 'triple combination' with conventional DMARDs (methotrexate plus sulfasalazine plus hydroxychloroquine) and the addition of several biologics to methotrexate improve chances to respond to treatment compared to methotrexate alone [36].

Because treatment with conventional DMARDs costs about 10-20 times less than the treatment with biologics, it is of interest to patients and payers whether the triple therapy is inferior or not to the

effective biologics.  The existing evidence cannot indicate any differences. Only Etarnecept was more efficacious than triple therapy (the odds of ACR50 response with triple therapy are on average 23% lower compared to Etanercept) while there is a lot of uncertainty to claim any important differences. Planning a further study to resolve this uncertainty (there is already one study available with 376 patients) would be of interest to the manufacturer or Etarnecept in order to promote it in for methotrexate'naive patients. Establishing a superiority for Etarncept might also be of interest to guideline developers and payers as it will clarify whether the added benefit of Etarnecept justifies its extra cost.

Figure 3 shows the power curves for the study planned as stand-alone experiment, based on conditional power for the head-to-head meta-analysis and the conditional power for NMA.  A two-arm trial randomizing 225 patients to MTX+Etarnecept and 225 patients to triple therapy should enable the updated NMA to detect an OR 0.77 with 80% power. The total sample sizes for 80% power for the trial alone or the head-to-head pairwise meta-analysis are 1,850 and 2,100 respectively.

**Figure 3 Power to detect an OR for ARC50 response of 0.77 between MTX+Etarnecept versus triple therapy as a function of the sample size. Power for a single RCT considered in isolation (dotted), conditional power for a random effects pairwise meta-analysis (dashed, two studies) and network meta-analysis (solid). Expected event rate in the triple therapy equal to the average observed in the network (49%). The network heterogeneity standard deviation is 0.04.**
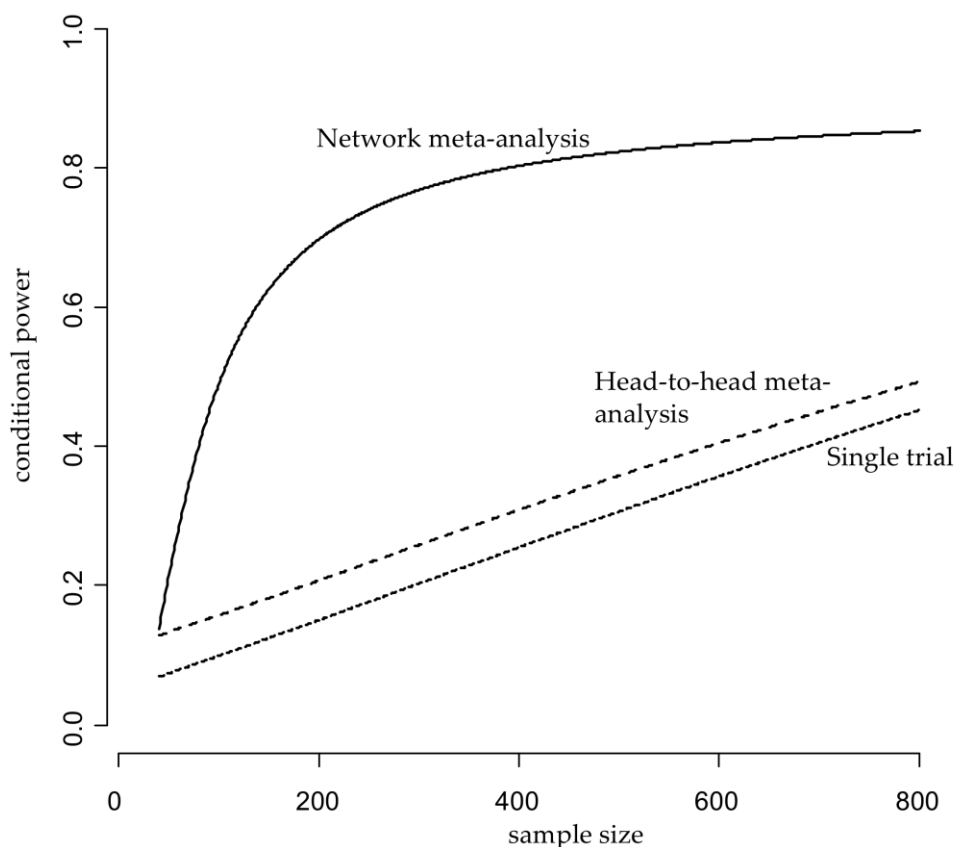
*4.3    Should we desing a new trial to help patients and clinicians choose between the most efficacious antispychotics?*

To take decision about optimal treatment, tolerability between the four most effective antipsychotics needs also to be taken into account. Clozapine and in particular Olanzapine have been associated with substantial weight gain, as also shown by Leucht et al. Consequently, patients who value their figure might want to prefer Amisulrpide and Risperidone over Clozapine unless the expected improvement in symptoms is substantial. Although Clozapine appears to decrease the symptoms more than Amisulpride (NMA SMD -0.22 (-0.41, -0.04)) or Risperidone (–0·32 (–0·47 to –0·16)), this advantage might not be enough to counterbalance the increase in weight. If  patients want an improvement in symptoms of at least SMD 0.20  in order to choose Clozapine over Amisulpride or Risperidone, then a new study needs to be planned to improve precision so that the 95% CI of the new NMA estimate would marginally exclude 0.20. However, this is impossible to achieve by adding a single new trial because of the heterogeneity present. Adding new data will shrink the confidence intervals up to a point beyond which no further power is gained. This is easy to show by considering that the variance of the summary estimate is a function of the sampling variances of the individual studies and the heterogeneity variance: new data will can decrease the total sampling variance to a very low number, but has no impact on heterogeneity which puts a "ceiling' on the improvement of precision, which remains constant throughout our calculations.

**Figure 4 Upper limit of the NMA OR for Clozapine versus Amisulpride and Clozapine over Risperidone when a new three-arm trial with 1:1:1 randomisation ratio is designed. The dashed line is the worthwhile effect and it is always included in the confidence interval. The heterogeneity is that estimated from the network ($\tau$=0.1)**

Figure 4 shows the upper confidence limits for each of the two NMA ORs as a function of the sample size of a new three-arm study, with expected effect sizes in the new studies equal to the NMA estimates. Adding more than 500 participants per arm does not make sense as the confidence intervals do not shrink any further. Possible solutions include re-estimation of the heterogeneity as evidence is being accumulated; restriction to populations and settings that might be more relevant and similar; e.g. younger patients and more recent studies; planning several small studies in order to understand heterogeneity better.

# 5 *How likely is it to adopt Actively Living Network Meta-Analysis in practice: results from a survey*

We conducted a survey to evaluate the acceptability and relevance of key methodological components of AL-NMA. Eligible participants were methodologists officially affiliated with one of the following a) a clinical trials unit located in Europe b) a pharmaceutical industry (only methodologists involved in clinical trials and HTA)  c) a European HTA agency d) the Cochrane Collaboration e) the World Health Organization (WHO, only methodologists involved in HTA, evidence synthesis, guideline development and WHO-funded trials). A detailed list of the various constituencies invited to participate in the survey is given in the Appendix.

The survey comprised two parts. The first part was about evidence synthesis and surveying the acceptability of NMA, how researchers decide about the conclusiveness of meta-analysis and maps the beliefs about the need of sequential monitoring in frequently updated evidence synthesis. The second part evaluated the degree to which meta-analysis is used to design a study and the acceptability and relevance of conditional power when design a new study. Participants were asked to answer one or both parts, depending on their expertise and affiliation.

## 5.1 *Description of the sample.*

In total 76 participants filled in the survey. The largest group represented was researchers were affiliated with a CTU unit (29 participants, 39%). The majority of the participants have been involved at least in one meta-analysis (62, 82%). Among them, 14 (23%) were very experienced (been involved with more than 20 meta-analyses). All participants knew what NMA is; and about half of them (36) have been involved with at least one NMA. The majority (66, 87%) have experience with RCTs and 43% are very experienced having worked in more than 20 RCTs. 44% of the participants have been involved with the production of clinical guidelines, 48% had some role in deciding about funding for clinical research and 37% have worked on producing methodological guidance for meta-analysis.

## 5.2 *Part I: Deciding upon conclusiveness in a meta-analysis*

The median number of participants answering this part was 60. About one in four participants (28%) believe that NMA should always be used, one in five that it should be considered in the absence of direct evidence only (19%), and about one in three that it should be used only when there are few direct studies (30%). However, in the absence of evidence of inconsistency a precise result

from NMA is trusted more than an imprecise result from a direct pairwise meta-analysis (45% versus 25%).

Assuming a meta-analysis with low heterogeneity and overall low risk of bias, researchers decide upon its conclusiveness by considering primarily the clinical importance (73%) and the statistical significance of the summary effect (58%). Only 7 participants (out of 55 who answered this question) answered that they consider the potential impact of future studies. Qualitative methods (possibly with involving stakeholders and often informal) were the most widely applied method to decide upon important harmful and beneficial outcomes when forming the conclusions of a review. One in five (11, 21%) use a decision analysis tool. To the question whether adjusting for multiple testing is needed in frequently updated evidence synthesis, replies were nearly equally split between 'yes' (31%), 'no' (40%) and 'I don't know' (29%).

## 5.3    *Part II: The role of meta-analysis in designing future studies*

The median number of participants answering questions in part II was 43. On a visual analogue scale, the average frequency of using results from meta-analysis in trials was 46%. The primary use of meta-analytic results, when applied, was to define the treatment difference sought and other parameters in sample size calculations (68%). Participants believed in their majority (67%) that using the results of meta-analysis to design a new study increases the chances of getting public finding. When answering an important health-policy related question participants believed that public funding should support by priority a NMA while the second most valued design was a three-arm trial comparing the two most promising treatments and standard treatment. The majority of participants were aware of the fact that a trial could be designed for conditional power of an existing meta-analysis (68%), but one in four (24%) said they are not willing to choose this design. Some of the reasons are presented in Box 2. One in four said they would consider this design (27%); the other two responded 'maybe' (49%).  The changing of paradigm in funders and research was pointed out as the major barrier towards adopting this trial design (49%). On the other hand, the majority of the participants (56%), when asked to reply under their capacity as citizens that support public trials, they said that priority should be given to trials designed using conditional power.

**Box 2 Sample reasons provided for not considering the conditional power of an updated meta-analysis when planning future trials**

---

• "Clinical trials are perceived as independent pieces of evidence. There would need to be a major shift by regulators, HTA bodies and physicians for companies to design trials in the context of meta-analyses"

• "Lots of examples where a large definitive trial has contradicted the results of a meta-analysis of smaller trials."

• "Usually the context in which I work is of trials supporting applications for a license. Regulators require each study to be 'significant' independently of others."

• "Wonder whether it would be convincing to authorities"

• "Any meta analysis is observational research"

• "Because when you finalize the trial, the meta-analysis will be outdated. Your study should be a stand alone trial."

• "Not enough faith in the homogeneity/comparability of the studies"

---

References

1. 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. doi:10.1371/journal.pmed.1001603.

2. Badgett RG, Vindhyal M, Stirnaman JT, Gibson CM, Halaby R. A Living Systematic Review of Nebulized Hypertonic Saline for Acute Bronchiolitis in Infants. *JAMA Pediatr*. 2015;169(8):788-789. doi:10.1001/jamapediatrics.2015.0681.

3. Charidimou A, Soo Y, Heo JH, Srikanth V, META-MICROBLEEDS Consortium. A call for researchers to join the META-MICROBLEEDS Consortium. *Lancet Neurol*. 2016;15(9):900. doi:10.1016/S1474-4422(16)30124-7.

4. Rahal AK, Badgett RG, Hoffman RM. Screening Coverage Needed to Reduce Mortality from Prostate Cancer: A Living Systematic Review. *PloS One*. 2016;11(4):e0153417. doi:10.1371/journal.pone.0153417.

5. Kanters, S, Ford, N, Druyts, E, Thorlund, K, Mills EJ, Bansback N. Use of network meta-analysis in clinical guidelines. *Bull World Health Organ*. 2016;94:782–784.

6. Brooks-Renney C, Costello S, Kusel J, Timm B. The use of mixed treatment comparisons in NICE technology appraisals. 2011.

7.   Eichler H-G, Thomson A, Eichler I, Schneeweiss S. Assessing the relative efficacy of new drugs: an emerging opportunity. *Nat Rev Drug Discov*. 2015;14(7):443-444. doi:10.1038/nrd4664.

8.   Créquit P, Trinquart L, Ravaud P. Live cumulative network meta-analysis: protocol for second-line treatments in advanced non-small-cell lung cancer with wild-type or unknown status for epidermal growth factor receptor. *BMJ Open*. 2016;6(8):e011841. doi:10.1136/bmjopen-2016-011841.

9.   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. doi:10.1186/s12916-016-0596-4.

10.  Efthimiou O, Debray TPA, van Valkenhoef G, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. January 2016. doi:10.1002/jrsm.1195.

11.  Garner P, Hopewell S, Chandler J, et al. When and how to update systematic reviews: consensus and checklist. *BMJ*. 2016;354:i3507.

12.  Takwoingi Y, Hopewell S, Tovey D, Sutton AJ. A multicomponent decision tool for prioritising the updating of systematic reviews. *BMJ*. 2013;347:f7191.

13.  Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet Lond Engl*. 2014;383(9912):156-165. doi:10.1016/S0140-6736(13)62229-1.

14.  Clarke M. Doing new research? Don't forget the old. *PLoS Med*. 2004;1(2):e35. doi:10.1371/journal.pmed.0010035.

15.  Cooper NJ, Jones DR, Sutton AJ. The use of systematic reviews when designing studies. *Clin Trials Lond Engl*. 2005;2(3):260-264.

16.  Jones AP, Conroy E, Williamson PR, Clarke M, Gamble C. The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. *BMC Med Res Methodol*. 2013;13(1):50. doi:10.1186/1471-2288-13-50.

17.  Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. *BMJ*. 2013;346:f1135.

18.  Goudie AC, Sutton AJ, Jones DR, Donald A. Empirical assessment suggests that existing evidence could be used more fully in designing randomized controlled trials. *J Clin Epidemiol*. 2010;63(9):983-991. doi:10.1016/j.jclinepi.2010.01.022.

19.  Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(2):80–97. doi:10.1002/jrsm.1037.

20.  Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. *MedDecisMaking*. 2013;33(1552-681X (Electronic)):597-606. doi:10.1177/0272989X13487604.

21. Puhan MA, Schünemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ*. 2014;349:g5630.

22. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. *PloS One*. 2014;9(7):e99682. doi:10.1371/journal.pone.0099682.

23. Naci H, van Valkenhoef G, Higgins JPT, Fleurence R, Ades AE. Evidence-Based Prescribing: Combining Network Meta-Analysis With Multicriteria Decision Analysis to Choose Among Multiple Drugs. *Circ Cardiovasc Qual Outcomes*. August 2014. doi:10.1161/CIRCOUTCOMES.114.000825.

24. Barrett B, Brown D, Mundt M, Brown R. Sufficiently Important Difference: Expanding the Framework of Clinical Significance. *Med Decis Making*. 2005;25(3):250-261. doi:10.1177/0272989X05276863.

25. Barrett B, Brown R, Mundt M, et al. Using Benefit Harm Tradeoffs to Estimate Sufficiently Important Difference: The Case of the Common Cold. *Med Decis Making*. 2005;25(1):47-55. doi:10.1177/0272989X04273147.

26. 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):2479-2500. doi:10.1002/sim.2704.

27. Roloff V, Higgins JPT, Sutton AJ. Planning future studies based on the conditional power of a meta-analysis. *Stat Med*. 2013;32(1):11-24. doi:10.1002/sim.5524.

28. Nikolakopoulou A, Mavridis D, Salanti G. Using conditional power of network meta-analysis (NMA) to inform the design of future clinical trials. *Biom J Biom Z*. 2014;56(6):973-990. doi:10.1002/bimj.201300216.

29. Nikolakopoulou A, Mavridis D, Salanti G. Planning future studies based on the precision of network meta-analysis results. *Stat Med*. August 2015. doi:10.1002/sim.6608.

30. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Med Decis Mak Int J Soc Med Decis Mak*. 2004;24(2):207-227. doi:10.1177/0272989X04263162.

31. Welton NJ, Ades AE, Caldwell DM, Peters TJ. Research prioritization based on expected value of partial perfect information: a case-study on interventions to increase uptake of breast cancer screening. *J R Stat Soc Ser A Stat Soc*. 2008;171(4):807-841. doi:10.1111/j.1467-985X.2008.00558.x.

32. Higgins JPT, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. *Stat Med*. 2011;30(9):903-921. doi:10.1002/sim.4088.

33. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol*. 2008;61(1):64-75. doi:10.1016/j.jclinepi.2007.03.013.

34. Nikolakopoulou A, Mavridis D, Egger M, Salanti G. Continuously updated network meta-analysis and statistical monitoring for timely decision-making. *Stat Methods Med Res*. September 2016. doi:10.1177/0962280216659896.

35. Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet Lond Engl*. 2013;382(9896):951-962. doi:10.1016/S0140-6736(13)60733-3.

36. Hazlewood GS, Barnabe C, Tomlinson G, Marshall D, Devoe D, Bombardier C. Methotrexate monotherapy and methotrexate combination therapy with traditional and biologic disease modifying antirheumatic drugs for rheumatoid arthritis: abridged Cochrane systematic review and network meta-analysis. *BMJ*. 2016;353:i1777.