The uptake of Bayesian methods in biomedical meta-analyses: a scoping review, 2005-2016

Robert L Grant

Honorary Research Fellow
Faculty of Health, Social Care & Education
Kingston and St George's, University of London

Managing Director
BayesCamp Ltd

robert@bayescamp.com

This is a post-print of a paper published in the Journal of Evidence Based Medicine (2019); 12(1): 69-75. BayesCamp Ltd makes this post-print available from robertgrantstats.co.uk, licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Abstract

Aim
Bayesian statistical methods can allow for more complete and accurate incorporation of evidence in meta-analyses. However, these methods remain under-utilised.
Methods
A scoping review was conducted to examine the proportion of biomedical meta-analyses that used Bayesian methods in the period 2005-16. The review also examined reproducibility of the work, the cited sources, the reasons for it, its success or failure, the type of model and prior distributions, and whether a mixture of Bayesian and frequentist methods were employed.
Results
We found that 1% of meta-analyses are Bayesian, and that the reporting and conduct of these was often poor. Data were published in 41% of analyses, and programs to run the analysis in 18%. Network meta-analysis was the most common reason, and became increasingly popular in recent years. In the majority of papers, models and distributions were either not reported or explained in such brief and ambiguous terms as to be uninformative.
Conclusions
More use needs to be made of Bayesian meta-analysis, and reporting needs to be improved. Greater awareness of these methods, and access to training in them, is essential.

Key words

evidence-based medicine
Bayesian
statistics

meta-analysis
systematic reviews

Background

Bayesian data analysis can be simply characterised as using probability as a mathematical tool to describe any unknown quantity.[1] While the more widely used frequentist methods restrict probability to statements about long (technically, infinite) runs of identical experiments, other sources of uncertainty such as the values of missing data or the extent of attrition bias can be explicitly included in a Bayesian analysis. Another advantage is that findings, such as the size of a treatment effect, can be described directly in probability terms (for instance, the probability that drug A has better outcomes on average than drug B, or the probability that diet X produces average weight loss greater than the minimum clinically important difference).

In meta-analysis, there are several challenges which can be addressed flexibly with Bayesian methods, including:

- network meta-analyses (NMAs), where we want to compare drugs A and B, but most trials have compared one of them with placebo (this is also sometimes called a mixed treatment or indirect treatment comparison)
- unreported statistics in the trials, such as standard deviations at endpoint, that are required for meta-analysis
- combining different statistics for the same outcome, such as one trial reporting mean change in an outcome measure, another reporting median change, and a third reporting the number of participants achieving a certain threshold

In all these problems, there is an unknown quantity which we have to guess at in order to combine the studies. Standard advice involves imputing a single value under some assumptions, for example making the missing endpoint standard deviation the same as the baseline standard deviation.[2] However, it is known that imputing a single number like this can distort subsequent analyses, by ignoring the uncertainty around it.[3]

Because Bayesian analysis allows probability distributions to be used to describe any unknown, we can incorporate many different sources of bias and variance into the meta-analysis that are not possible in frequentist analysis. Sometimes, a frequentist analysis will be followed by some sensitivity ("what-if") analysis of biases, and sometimes by a qualitative statement of limitations, but the Bayesian approach is reproducible and open to examination and critique. These are some of the common problems in meta-analysis that could be included:

- Network meta-analysis allows papers comparing A and B (direct comparisons) to be combined alongside some that compare A with X, and others that compare B and X (indirect comparisons)
- Structural equation modelling allows us to estimate latent variables (such as impairment after stroke) that are imperfectly measured in various different outcome measures. Thus, papers with related but not identical outcomes can be combined, without having to make gross assumptions to convert the outcome values.
- Prior information can be incorporated, such as expert opinion or previous studies in vitro or in animal models.
- Randomised and observational studies can be combined, allowing for the additional uncertainty or bias arising from confounding in the observational data
- Papers that are of poor methodological quality can be included with extra uncertainty or bias around their statistics, rather than having to exclude them entirely

- Papers that report dichotomised outcome scales (% responders to treatment) can be incorporated with those that report mean outcomes on the same scale.

Bayesian methods were regarded as under-utilised in meta-analysis in a 2000 review.[4] The reasons for using them are still commonly given as concerns or limitations in published frequentist meta-analyses, so it would seem that much of this work would benefit from wider adoption of Bayesian methods. The software to carry out Bayesian meta-analysis has become more flexible, robust and efficient, and there are several free packages available. A more recent review of three high-profile multidisciplinary journals (Nature, Science and Proceedings of the National Academies of Science) in 2017 found 2.5% of papers using Bayesian methods.[5]

However, there are barriers for clinician-researchers: their statistical training generally does not include Bayesian models, statisticians are in short supply, and Bayesian analysis often involves writing out a model for the data and parameters of interest in code like a programming language, which can be daunting.

The adoption of network meta-analyses (NMA) has also accelerated since 2009, and a 2014 review found that at least 76% of NMAs used Bayesian software.[6] Seven items were identified in 2005 guidelines ("ROBUST") for reporting Bayesian analyses,[7] and a review of the reporting of NMAs also proposed quality criteria specific to that type of MA.[8]

The Cochrane Collaboration handbook, which is widely used by biomedical researchers as a guide to good practice, first included advice on Bayesian meta-analysis in February 2008.[9] Also in 2008, Sutton & Higgins published a review paper called "Recent Developments in Meta-analysis", which promoted Bayesian methods.[10] Taking these as landmarks that popularised Bayesian approaches for a wider audience of researchers, we reviewed meta-analyses published from 2005 to 2015, anticipating a particular adoption of Bayesian methods from 2008 onwards.

Methods

This is a scoping review,[11] which aims to determine what methods have been used and not what the conclusions of the individual meta-analyses were; we also examined "ancestor" publications, cited as explanation or justification of the methods employed. The papers under consideration were obtained by systematic literature search. The focus of this review was the uptake of Bayesian methods in meta-analysis of biomedical / healthcare studies (translational or clinical, not basic, research). The following research questions were chosen, based on contemporary concerns about reproducible research [12] and in discussion with expert "critical friends":

1. What proportion of meta-analyses in biomedical topics have used Bayesian methods; is there evidence of this proportion increasing or decreasing over time?
2. How reproducible were the meta-analyses: was code provided; were data provided; was the software version named; what software was used, and was it open-source?
3. What ancestors[11] were cited for Bayesian MAs?
4. What was the reason for using a Bayesian approach?
5. Did the authors report the Bayesian aspects of the MA as successful?

6. What sort of probabilistic model was employed?
7. What sort of prior distributions were employed?
8. Did the MA include both frequentist and Bayesian versions of the same analysis; if so, did the authors justify this?

A search was conducted in August 2015 (and updated in January 2018) on the MEDLINE and CINAHL databases for peer-reviewed English-language papers on human subjects with the term "meta-analysis" in the abstract, published since 1 January 2005. The EMBASE database was considered but not included because it contains all journals in MEDLINE with additional sources in pharmacology and drug development, which was felt unlikely to contain meta-analyses that were not in MEDLINE. The indexed publication types "review" or "systematic review" were required in CINAHL, and "meta-analysis" was required in MEDLINE. These searches were then re-run, additionally requiring the word "Bayes" or " credible interval" in the abstract, to identify the subset of Bayesian papers. Credible intervals are a Bayesian analogue to confidence intervals. Duplicates between the databases were removed.

The full text versions of the Bayesian papers up to end of 2014 were screened manually to address the research questions. The 2015 and 2016 publications obtained in the update search were only included to answer question 1 and a simplified version of question 4: whether it is a network meta-analysis or not. Papers which were revealed to be methodological, or protocol papers, or otherwise ineligible, were excluded. The non-Bayesian papers were not examined further. Figure 1 shows the numbers of papers at each step. This review was not funded, so double-extraction was not possible.

Because statistical methods are not indexed in the databases, papers using specific methods can only be identified by searching for specific terms in the abstracts and titles. It is possible that a Bayesian meta-analysis might not mention “Bayes” or “credible interval”. To estimate the proportion of apparently frequentist papers that are actually Bayesian, we selected a random sample of apparently frequentist NMAs , and examined them in full text to count how many were actually Bayesian. They were identified by the terms “network meta” or “indirect” and then screened manually. NMAs were chosen because they were the most common form of Bayesian meta-analysis, and are easily identified in the abstracts. The sample size for this exercise (81) was chosen to obtain a posterior beta distribution, assuming a prevalence of 40%, with credible interval not extending outside 30-50%. A prior of beta(2,3) was used for this, corresponding to a weak expectation of some missed Bayesian papers, with mean prevalence of 40%. To obtain 81 papers, a list of the numbers identifying all of the apparently frequentist NMAs was sorted into random order. Each of these was then examined in turn, setting aside those that were not NMAs, until 81 had been selected.

All automated text searching and processing was done by programs in R (version 3.3.1) or C++ (using the C++11 standard), and full details are given online.[13] Percentages and counts of papers are presented below; credible intervals or confidence intervals are not calculated around these, because MEDLINE and CINAHL represent a large proportion of biomedical publications of high methodological rigor, suggesting that the papers obtained are close to being the entire population of biomedical meta-analyses.

Findings

The process of refining the search results to arrive at the counts of MAs and Bayesian MAs is shown in Figure 1. The first search yielded 35,260 MA papers and 312 Bayesian MA papers. The 312 were examined in full and answer questions 2-8. Of this search, 33,911 MA papers and 286 Bayesian MA papers were in the period 2005-14 and they are combined with the results of the update search to answer question 1 and a simplified version of question 4. The update search yielded 17,259 MA papers and 226 Bayesian MA papers.

There were two grey areas for classification. The term "meta-analysis" was occasionally used in genomic studies to indicate a different kind of analysis (stratified with random effects). The line between individual patient data meta-analysis and multilevel modelling in secondary analyses such as prevalence studies is not clear-cut.

There were 432 papers that appeared to be frequentist NMAs. We reviewed a random selection of 104 in full-text. Of these, 23 turned out not to be NMAs, and 21 out of 81 were Bayesian (26%). Combined with the prior distribution used in the sample size calculation (beta(2,3)), this gives a 95% credible interval for the prevalence of Bayesian NMAs that do not declare themselves as such in the abstract extending from 17% to 36%. Extrapolating these proportions to all 432 papers, we might expect 112 to have gone undetected over the whole time period of the search. We believe the number of other forms of meta-analysis which we incorrectly classified as frequentist to be negligible, as the frequentist software tools for these are so widely used, and so it was not feasible to examine a large number of papers to estimate their prevalence.

Question 1: What proportion of meta-analyses in biomedical topics have used Bayesian methods; is there evidence of this proportion increasing or decreasing over time?

Over the whole period of the review including the update search, 512/51,170 papers were Bayesian, a prevalence of 1.0%. Figure 2 shows the annual count of meta-analyses, Figure 3 the percentage of Bayesian meta-analyses, and Table 1 the counts. Although the number of meta-analyses increased eight-fold in ten years, there was only a slight increase in the percentage of these which used Bayesian methods, and 2015 and 2016 had the highest of these percentages.

Question 2: How reproducible were the meta-analyses: was code provided; were data provided; was the software version named; what software was used, and was it open-source?

Many papers relegated details of models to an online appendix. This sometimes introduced practical barriers to reproducibility: some were paywalled even if the paper was not, sometimes the link was broken, and some were obtainable but not in a readily machine-readable form (often Word files or PDF). Code was provided in 55/312 (18%) and data in 128/312 (41%), with a further 18 where it is not clear if all data are present. Often the data are available only in forest plots. The software was named by 279/312 (89%), and the version was given by 144/279 (52%) of these. BUGS was the most popular with 230/312 papers (74%), 212 of which were WinBUGS, followed by: JAGS (10), GeMTC (8), R (no named package) (7), ADDIS (5), PyMC (4), SAS (3), Stata (3), Microsoft Excel (3), Lotus 1-2-3 (3), R 'HSROC' package (2), bespoke FORTRAN (2), and 1 each of AgenaRisk, bespoke C, DisMod-MR, FAST*PRO, MANTRA, MMM, Stan, and R packages 'MCMCglmm' and

'inla'. In 40 papers, WinBUGS 1.4 was given as the version, but there were actually three versions from 1.4.1-1.4.3. The software named in 4 papers seemed erroneous: either it did not have Bayesian capability or version numbers were given that did not exist.

Question 3: What ancestors were cited for Bayesian MAs?

The most common ancestors were related to NMAs: Lu & Ades 2004[14] (one of the earliest papers on NMA, written from a theoretical perspective for a statistics journal) were cited by 59 papers, Caldwell *et al* 2005[15] (which introduced NMAs to a medical readership) by 27, NICE DSU[16] (a report giving practical guidance on implementing NMAs, including BUGS code) by 24, and Salanti *et al* 2008[17] (a review of NMA methodology that introduced some theorems for when they will work effectively) by 16. The most common of the more general ancestors were: Spiegelhalter, Abrams & Myles 2004[18] (a relatively early book on Bayesian methods in clinical trials: cited by 19 papers), Warn, Thompson, Spiegelhalter 2002[19] (a theoretical statistics paper examining measures for binary outcomes in Bayesian MA: 12 papers), and Smith, Spiegelhalter & Thomas 1995[20] (a review of different models in Bayesian MA: 11 papers). Ades, Sculpher *et al* 2006[21], a health economics paper, was cited by 11 papers. All other ancestors were cited by fewer than ten papers.

Question 4: What was the reason for using a Bayesian approach?

NMA was the most common problem that the Bayesian analysis sought to solve, in 151/312 papers (48%). Including the update search for 2015-16, this rose further to 288/512 (56%). If we consider the extrapolated figure of 112 missing Bayesian NMAs that do not name themselves as such, this could be as high as 400/624 (64%). The change over time is shown in Figure 4. Other common problems were classified as diagnostic meta-analysis (22/312: 7%), meta-regression for risk factors in observational data (16/312: 5%), and genetic or genomic models (14/312: 4%). In most other studies, no particular justification was given.

Question 5: Did the authors report the Bayesian aspects of the MA as successful?

It was extremely rare to find a paper that declared that its methods had not achieved their goals (2/312: 0.6%) or had mixed success (13/312: 4%).

Question 6: What sort of probabilistic model was employed?

It was not possible to classify the great majority of papers with any confidence as information was so lacking. In many papers, models are only described as 'hierarchical' or 'adjusting for study-level covariates' without further detail. Only a few papers adopted truly bespoke methods, and some declared that they had adopted previously published code.

Question 7: What sort of prior distributions were employed?

Many papers (86/312: 28%) did not declare any particular prior distribution other than general terms like "vague", 39/312 (13%) gave incomplete or contradictory information, and

102/312 (33%) did not mention priors at all. The great majority of those that gave complete information used diffuse or flat priors (68/85: 80%), while 7 (8%) were weakly informative, 3 (4%) subjective, and none had empirical priors. The terms "diffuse", "flat", "non-informative", "uninformative", "loosely informative" and "minimally informative" were all employed by various papers to indicate diffuse priors.

Question 8: Did the MA include both frequentist and Bayesian versions of the same analysis; if so, did the authors justify this?

Both frequentist and Bayesian analyses were undertaken in 130/307 (42%) of papers, with five unclear. Many of the NMAs (64/151: 42%) used frequentist analyses for direct effects and then Bayesian analyses for the indirect effects, which is common practice but it is hard for readers to understand which to take at face value. There is also a danger of inflated type 1 errors if the more exciting set of results is used to draw overall conclusions. Of the other papers that mixed approaches, it was most common for there to be no justification (45/66: 68%).

Only 8 papers gave sound justifications, an example of which is a meta-analysis of diagnostic studies in depression screening, which considered whether two screening questions were superior to a single question[22]. This paper used a common frequentist calculation for diagnostic studies, and then went on to construct a set of curves relating pre-test and post-test probabilities of depression for the various sets of questions under consideration. This latter part of the analysis required a Bayesian approach, and was justified as assisting communication of the results (diagnostic studies are a challenge for communication to clinicians).

The remaining 13 meta-analyses were unclear in some way in justifying a mix of analytical methods. For example, in a network meta-analysis of treatments for osteoarthritis pain[23], "Frequentist and Bayesian methods were used […] The frequentist meta-analysis using Bucher indirect comparisons was chosen because it reports traditional statistical measures, whereas the Bayesian network meta-analysis allows for inclusion of both direct and indirect information in a single step." This leaves readers uncertain about how to balance the presumed communication advantage of "traditional statistical measures" with the presumed statistical advantage of "a single step".

In many cases, the Bayesian analysis was used for some aspect that could be done in frequentist fashion, such as adjusting for study-level covariates, though not by preset functions in common software.

Discussion

Despite many possible benefits, Bayesian methods are employed in only about one percent of published meta-analyses, which has not obviously changed over ten years. In general, frequentist methods are used unless it is impossible to apply them to the problem at hand, or if an off-the-shelf Bayesian model can be adapted easily. This leads to some inconsistent mixtures of methods, which is potentially confusing and we suggest should be avoided. Some papers are coy about the Bayesian aspects, for instance by not mentioning software but saying that Markov Chain Monte Carlo simulation was used to estimate treatment effects.

Authors may feel it expedient to play down more unusual methods in order to avoid unwarranted criticism from reviewers.

Reproducibility should be prioritised by authors, reviewers and journals, with code and data provided online.[12] The ROBUST guidelines for reporting Bayesian analyses recommend giving details of the statistical model and the analytical technique. Meta-analyses generally use previously published statistics, so there is no reason not to publish data. However, there may be constraints on authors of individual patient data meta-analysis that prevent this. Many Bayesian papers mentioned single imputations of missing statistics like correlations or endpoint standard deviations; these should be incorporated in the Bayesian model, otherwise the implication is that they are precisely known. We did not attempt to count these as they are likely to be omitted from methods in many papers.

The great majority of papers used non-informative priors. We do not know what thought went into those decisions, if any, but in many cases a weakly informative prior would more accurately reflect clinical and statistical knowledge and could improve computational speed and stability without affecting the results.[1] The ROBUST guidelines recommend specifying and justifying the prior distribution, but only 27% of papers in this review specified them and fewer justified them. Very few also reported sensitivity analyses with different priors, which was another ROBUST recommendation, though this was not a specific question in this review.

This review has been limited by the inclusion of particular journals in the databases, and changes in that over time. Bayesian methods are not indexed in the databases, and not necessarily mentioned in the abstract, so we are likely to have missed some. Without specific funding, we have not been able to perform double-extraction from the papers. Also, conducting qualitative research such as interviewing authors to understand the barriers to adoption of Bayesian methods could be fruitful in future research.

Recommendations

Any meta-analysis should consider using a Bayesian model, as this can help with many problems that do not fit into models implicit in the common frequentist calculations. If Bayesian methods are to be used, then there should not be a mixture of frequentist and Bayesian, as this can confuse readers. It can be difficult to include all details of code, data, priors and sensitivity analyses in a journal paper, so authors should seek to publish online appendices for this purpose. These typically fall outside any copyright agreement regarding the paper, and so, to future-proof availability, can be replicated on the websites of publishers, institutions and the individual authors. Meta-analysts should consider software other than BUGS (for example, PyMC3 and Stan), which could offer advantages in speed and stability of calculations.

Acknowledgements:

This work was not funded. I am grateful to critical friends [names and affiliations have been withheld for double-blind review] for suggestions on an early draft.

References

1. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*, 3rd edn. Boca Raton, FL, USA: CRC Press, 2013.

2. Carpenter JR, Kenward MG. *Multiple Imputation And Its Application*. Chichester, UK: Wiley, 2013.

3. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

4. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technol Assess* 2000; 4(38).

5. Cristea IA, Ioannidis JPA. P values in display items are ubiquitous and almost invariably significant: a survey of top science journls. *PLoS One* 2018; 13: e0197440.

6. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidem* 2014; 67: 138-43.

7. Sunga L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *J Clin Epidem* 2005; 58: 261-8.

8. Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect Comparisons: A review of reporting and methodological quality. *PLoS One* 2010; 5: e11054.

9. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.0, section 16.8.1. The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.

10. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Stat Med* 2008; 27: 625–50.

11. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005; 8: 19-32.

12. *Science as an open enterprise: open data for open science*. The Royal Society, 2012. http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.

13. https://github.com/robertgrant/bayesian-dichotomised-meta-analysis

14. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23: 3105-24.

15. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Br Med J* 2005; 331: 897.

16. Dias S, Welton NJ, Sutton AJ, Ades AE. *A generalized linear modelling framework for pairwise and network meta-analysis of randomized controlled trials*. National Institute for Health and Clinical Excellence, Decision Support Unit. http://nicedsu.org.uk/wp-content/uploads/2017/05/TSD2-General-meta-analysis-corrected-2Sep2016v2.pdf

17. Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17: 279-301.

18. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, UK: Wiley, 2004.

19. Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Stat Med* 2002; 21: 1601-23.

20. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995; 14: 2685-99.

21. Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, Lu G. Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics* 2006; 24: 1-19.

22. Mitchell AJ. Are one or two simple questions sufficient to detect depression in cancer and palliative care? A Bayesian meta-analysis. *Br J Cancer* 2008; 98: 1934-43.

23. Myers J, Wielage RC, Han B, Price K, Gahn J, Paget M-A, Happich M. The efficacy of duloxetine, non-steroidal anti-inflammatory drugs, and opioids in osteoarthritis: a systematic literature review and meta-analysis. *BMC Musculoskeletal Disorders* 2014; 15: 76.

Table 1: Counts of meta-analysis papers

Year	Meta-analyses	Bayesian meta-analyses	Bayesian network meta-analyses
2005	1136	5	0
2006	1360	12	1
2007	1599	12	2
2008	1862	12	1
2009	2283	25	5
2010	2832	18	6
2011	3711	24	12
2012	5225	49	20
2013	6668	60	33
2014	7235	69	53
2015	8776	113	72
2016	8483	113	83

Figures:

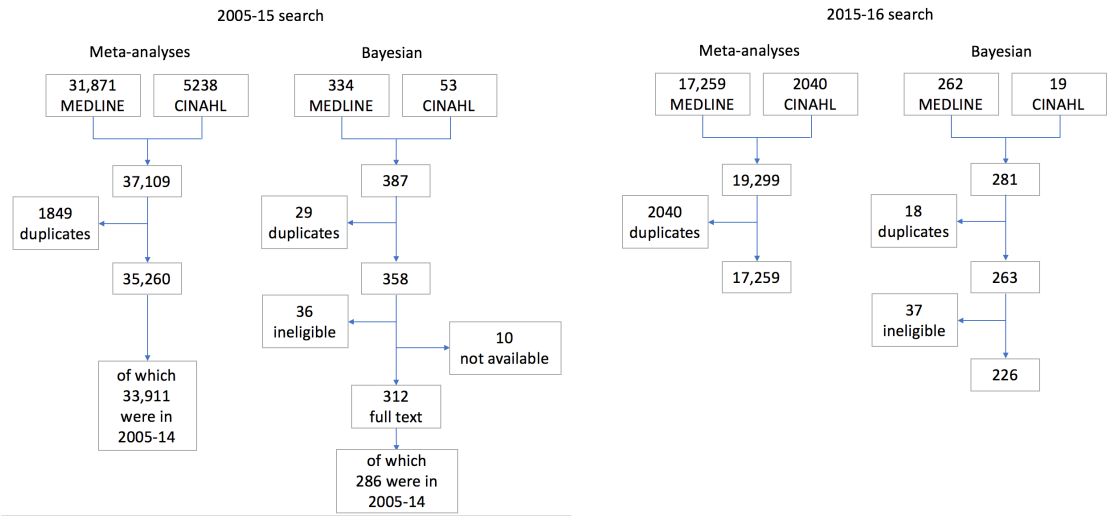

Figure 1: flowchart of searches and selection of abstract and full-text papers

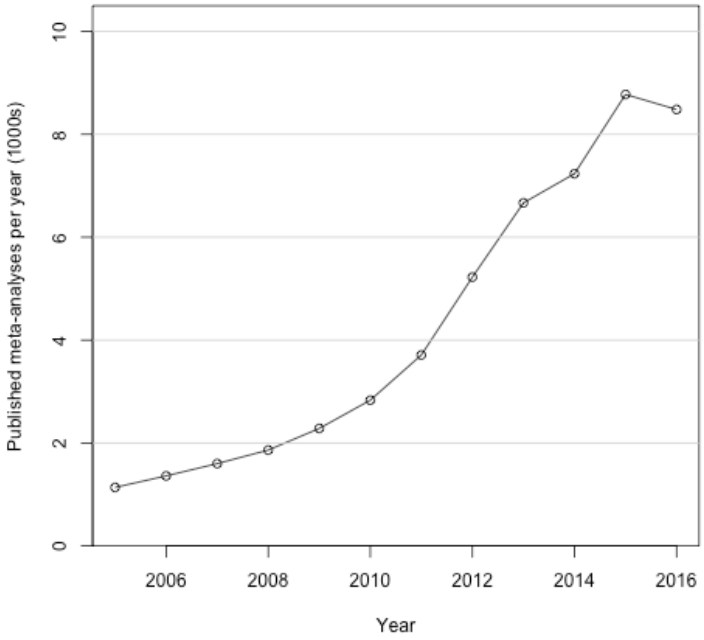

Figure 2: the number of biomedical meta-analyses has increased rapidly

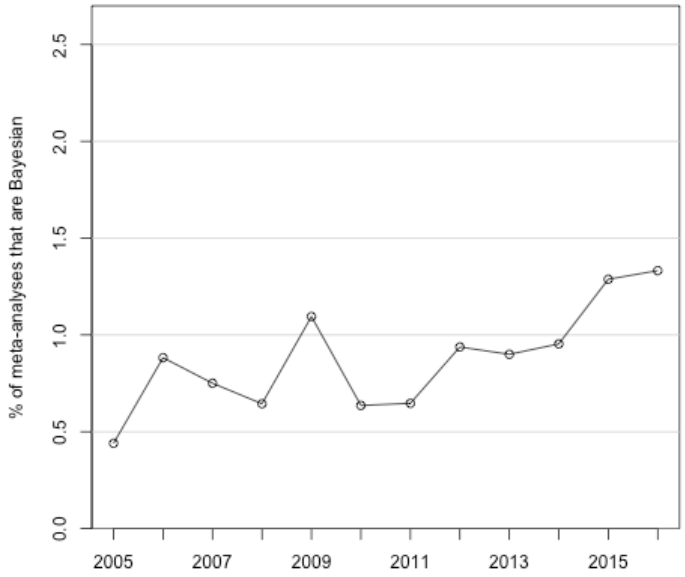

Figure 3: the proportion of meta-analyses that are Bayesian has increased slightly but remains low

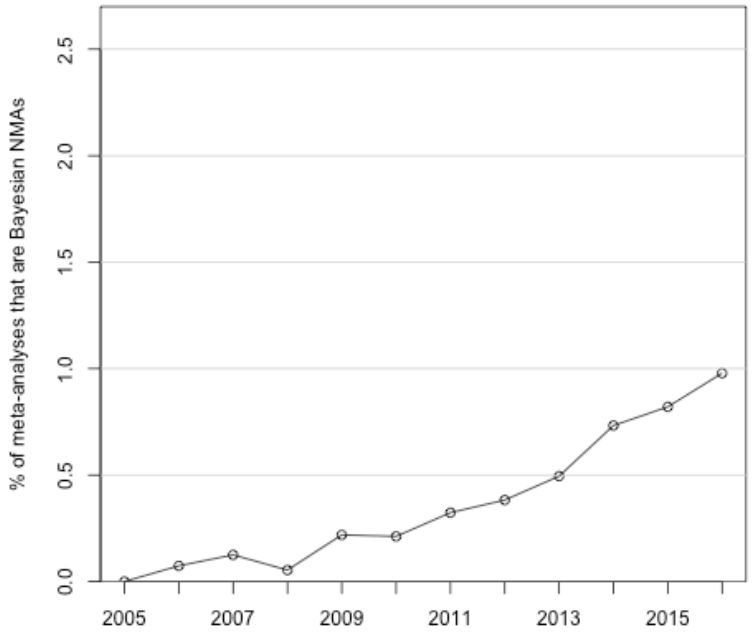

Figure 4: Bayesian network meta-analysis has rapidly increased in popularity, now making up most of the Bayesian meta-analyses