NETWORK META-ANALYSIS USING DATA FROM PUBLISHED TRIALS
AND DATA FROM THE FOOD AND DRUG ADMINISTRATION
MEDICAL REVIEWS: A CASE EXAMPLE OF FIRST LINE MEDICATIONS
FOR OPEN ANGLE GLAUCOMA

by
Qiyuan Shi

A thesis submitted to Johns Hopkins University in conformity with the
requirements for the degree of Master of Health Science

Baltimore, Maryland
May, 2015
Abstracts

Background

Network meta-analysis (NMA) is a method that incorporates both direct evidence within trials and indirect evidence across trials. The impact of different data sources on NMAs is not widely examined.

Objectives

The objective of this study was to conduct and compare the results from NMAs using clinical trial data available in the Food and Drug Administration (FDA) medical reviews (FDA trials) and those published in the medical literature (published trials) on first line medications for open angle glaucoma.

Methods

We searched the Cochrane Central Register of Controlled Trials, MEDLINE, EMBASE in March 2014 and the US Food and Drug Administration’s website in April 2014 for randomized, parallel group trials on the first line topical medications for primary open-angle glaucoma and ocular hypertension. Eligible trials had to compare a single active treatment with no treatment/placebo or another single active topical medical treatment. Two individuals independently assessed trial eligibility, abstracted data, and assessed the risk of bias. Using the FDA trials and the published trials, we performed pair-wise meta-analyses and Bayesian network meta-analyses on intraocular pressure at 3 months, our pre-specified primary outcome. We compared the results from these analyses.

Results
We included 16 FDA trials and 105 published trials. The network based on FDA trials had 10 nodes (nine active drugs and placebo/vehicle) with a total sample size of 6183. The network based on published trials had 14 nodes (13 active drugs and placebo/vehicle) with a total sample size of 16898. There were 9 common treatments between the FDA trials and the published trials in the network meta-analysis, resulting in 36 comparisons. We estimated mean difference in IOP of 36 pairs of these treatments. The median relative difference of the two networks was 14% (interquartile range: 7% to 42%). The relative differences were greater than 25% in 14 pairs. Point estimates from two pairs showed opposite directions. Generally, the results of the published trials had better precision than the results of the FDA trials because of the larger sample size of the published trials. In terms of relative rankings, bimatoprost, travoprost, and latanoprost are the best ranking drugs regardless of the data source, although 4/9 (44%) drugs had different ranking.

**Conclusions**

Systematic reviewers should consider FDA medical reviews because of the amount of the information they can provide. The data source did not seem to change the inference of the relative effectiveness of most of the drugs studied. The relative rankings changed for the middle-ranked drugs but not for the top-ranked drugs. Whether reporting bias has a role in the differences we observed needs to be further evaluated.

**Advisor:** Tianjing Li, MD, MHS, PhD

**Second reader:** Mark L. Van Natta, MHS
Acknowledgements

Before proceeding to the contents of this study, I would like to make some acknowledgements to thank the people who contributed to this study or gave me help of all kinds. Foremost, I would like to express my sincere gratitude to my advisor Dr. Tiangjing Li for the continuous support of this study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. Second, I want to thank Mark L. Van Natta for being my second reader and for his great advice and instructions. I would also like to thank my dear colleague Benjamin Rouse for his support and help.

I would like to thank Hwanhee Hong, Sueko Ng, Nan Guo Cesar Ugarte-Gil, Manuele Michelessi, Lijuan Zeng, and Kexin Jin for their help with screening, data abstraction, and analysis.

Finally I would like to thank my parents for their support and love for all the time.
Table of Contents

1 Introduction ........................................................................................................................................... 1

2 Methods .............................................................................................................................................. 5
   2.1 Eligibility Criteria .......................................................................................................................... 5
   2.2 Retrieving Medical Reviews from the FDA Website ................................................................. 5
   2.3 Retrieving Published Trials ......................................................................................................... 6
   2.4 Data Abstraction and Management ............................................................................................. 6
   2.5 Outcomes ..................................................................................................................................... 6
   2.6 Pairwise and Network Meta-analysis ......................................................................................... 6
   2.7 Evaluation of Clinical and Methodological Heterogeneity ....................................................... 8
   2.8 Evaluation of Statistical Heterogeneity and Inconsistency ....................................................... 8
   2.9 Comparison between Results from Different Sources ............................................................... 9

3 Results .............................................................................................................................................. 11
   3.1 Identification of Trials .................................................................................................................. 11
   3.2 Analysis of FDA trials .................................................................................................................. 11
   3.3 Analysis of Published Trials ...................................................................................................... 12
   3.4 Comparison between FDA Trials and Published Trials ............................................................ 14
   3.5 Heterogeneity ............................................................................................................................. 16
   3.6 Inconsistency ............................................................................................................................... 17

4 Discussions .................................................................................................................................... 18

5 References: ...................................................................................................................................... 22

6 Appendices ...................................................................................................................................... 25

7 Tables .............................................................................................................................................. 31

Table 1. Summary estimates for intraocular pressure at 3 months derived from pair-wise meta-analysis based on 13 direct comparisons from 16 FDA trials........31
Table 2. Summary estimates for intraocular pressure at 3 months derived from network meta-analysis of 16 FDA trials ........................................32
Table 3. Summary estimates for intraocular pressure at 3 months derived from pair-wise meta-analysis based on 36 direct comparisons from 105 published trials33
Table 4. Summary estimates for intraocular pressure at 3 months derived from network meta-analysis of 105 published trials..........................................................34
Table 5. Comparisons between the results of direct pairwise meta-analyses based on FDA trials and published trials.................................................................35

8 Figures ........................................................................................................................................36

Figure 1. Identification of trials.................................................................................................36
Figure 2. Network graphs........................................................................................................37
Figure 3. Ranking probabilities for any drug at any position in the FDA trials network .................................................................................................................38
Figure 4. Ranking probabilities for any drug at any position in the published trials network.................................................................................................................39
Figure 5. Scatterplot of relative effect size between two treatments based on both FDA and published trials.................................................................................40
Figure 6. Relative effect size based on FDA trials and published trials ............................41
Figure 7. Relative effect size between active drugs and placebo/timolol based on FDA trials and published trials........................................................................42
Figure 8. Percentage ‘SUCRA’ of drugs in the two networks .............................................43
1 Introduction

Since the concept of evidence-base medicine was introduced, systematic reviews are often regarded as the highest-level of evidence [1, 2]. Because systematic reviews use ‘explicit and systematic methods to minimize bias’, they are thought to provide more reliable findings than individual research studies from which conclusions and decision can be made [3]. However, a typical systematic review focuses on the synthesis of direct, pair-wise comparisons that are available from at least one study. If no such study is available, for example, if no study has examined the comparative effectiveness of treatment A and treatment B, it becomes a limitation for conventional systematic reviews. Such needs lead to the development of network meta-analysis (NMA).

NMA, as is indicated by its name, aims to build a network where nodes represent interventions while the edges represent direct comparisons (that is available from at least one included study) [4]. If two interventions A and B have one intermediate comparator C, A and B can be compared indirectly generating indirect evidence. By weighting both direct and indirect evidence, a mixed effect can be estimated. Thus, NMAs can accomplish the task of “all-way” comparisons with more information [4]. A unique feature of NMAs is that it also can generate ranking probabilities of all the interventions in terms of their efficacy and/or potential for harm[5]. Conclusions on the comparative effectiveness of all interventions included in a network could be drawn based on both the estimates of relative effects and ranking probabilities.
Systematic reviews are often subject to reporting bias, that is “systematic differences between reported and unreported findings” [3]. Two common types of reporting bias are publication bias and selective outcome reporting bias. Publication bias refers to whether a study is published is dependent on the direction and magnitude of the results; selective outcome reporting bias refers to bias in published studies due to partial or distorted reporting of the outcomes [3]. Numerous studies have shown that reporting bias is prevalent in clinical trials and has become a threat to the validity of systematic reviews [6-8].

Because NMA is a newly developed methodology, the impact of reporting bias on NMA is not well understood. One study by Trinquart et al., examining an antidepressants network based on clinical trial data obtained from the Food and Drug Administration (FDA), found that reporting bias affecting one drug could change the relative rankings of drugs in the whole network [9]. In this study, medical reviews of trials prepared by the FDA (we will use FDA trials for simplicity) were considered as the reference (‘truth’). After conducting a sensitivity analysis on each drug, the authors found that if the FDA trials were partially reported in the scientific journals, the relative rankings of drugs based on the incomplete data set would differ from the ‘truth’, where all FDA trials were used.

In another study, Song et al. found that if all placebo-controlled trials were biased in
favor of “new” or “sponsored” treatments to the same extent, the ‘indirect comparison will counterbalance such bias’ [10]. This finding suggests that NMAs might be more robust to reporting bias under certain circumstances than pair-wise meta-analysis. The same phenomenon was observed in a study of citalopram and escitalopram in which the indirect evidence shows no evidence of difference while the direct evidence favors the new, sponsored drug (in this case escitalopram)[11]. The authors speculated that there was bias in the direct evidence.

Investigating reporting bias requires unique methods and sources of data. Registry data such as data submitted to the regulatory agencies for gaining marketing approval have become a popular source. Using the FDA in the United States as an example, manufactures must submit clinical trial data including protocols and raw data from the trials to the FDA before a drug can be approved. Scientists at the FDA review these data and prepare reports that summarize their evaluation and decisions. Most of these documents prepared by the FDA, especially the more recent ones are posted on the FDA’s website for public viewing and use (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/).

These same trials sponsored by drug companies also may be published in scientific journals, although not all of them are published. Trials with negative and null results are less likely to be published than trials with positive results [12]. Studies also have shown that the trial data submitted to the FDA and those published in the literature


sometimes disagree in sample size, effect size, and in the analysis plan [13].

To further investigate the impact of reporting bias on NMA, in particular, the value of using FDA trials for NMA, we need to know the effect estimates and rankings of interventions based on (1) the FDA trials alone; (2) published trials alone; (3) all trials.

For the purpose of this project, we focused on analyses (1) and (2). We built upon an ongoing systematic review and network meta-analysis [14] on the comparative effectiveness and safety of first-line intraocular pressure (IOP) lowering drugs for patients with primary open angle glaucoma (POAG) or ocular hypertension (OH). Our study is a case example of examining the impact of reporting bias on NMA.

The objectives of this study were:

1. To identify and abstract data from FDA reviews of trials on first-line IOP lowering drugs for patients with POAG or OH.

2. To conduct pairwise meta-analyses and NMAs using (1) data abstracted from the FDA trials; (2) data abstracted from all published trials.

3. To compare the results from the two analyses above and evaluate the impact of different data sources on the findings from NMA.
2 Methods

2.1 Eligibility Criteria

Trials were eligible for our NMA if they (1) recruited more than 60% patients diagnosed with POAG or OH, using any definition specified in the trial; (2) compared a first-line topical IOP lowering drug to placebo/no treatment or another IOP lowering drug; (3) randomized controlled trial (RCT) that used a parallel design.

Trials were excluded if it (1) enrolled fewer than 10 patients in any study arm; (2) evaluated combination medical therapies for POAG or OH; (3) had a follow-up duration shorter than 28 days after randomization.

2.2 Retrieving Medical Reviews from the FDA Website

As described previously by Turner et al.[15], FDA medical reviews of approved drugs are available from the FDA website (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/). We searched drug names (active ingredients) and downloaded medical reviews in electronic form, including ‘Approval History’, ‘Letters’, ‘Reviews’, and ‘Related Documents’. For those medical reviews that are not available from the FDA website, which are usually medical reviews for drugs approved before 1996, we filed a request to FDA. We did not receive any response from the FDA by the timeline allowed for this project. We proceeded with medical reviews available for download from the FDA website. The list of drugs we searched is available in Appendix 1.
2.3 Retrieving Published Trials

We used trials identified from an ongoing systematic review and NMA [14]. In brief, Li et al. searched MEDLINE, EMBASE and Cochrane Register of Controlled Trials (CENTRAL) for eligible trials up to March 2014. The search strategies are available in Appendix 1.

2.4 Data Abstraction and Management

Two individuals working independently abstracted data from included trials into an online database developed in the Systematic Review Data Repository (http://srdr.ahrq.gov/) [16]. We adjudicated discrepancies through discussion.

2.5 Outcomes

IOP was typically used as the primary outcome in both published trials and FDA trials because it is the endpoint upon which the hypotensive glaucoma drugs is approved [17]. Thus, for all of the analyses in this study, we used mean IOP at 3 months as the outcome measurement. If the 3 months IOP data were not available, we used the time point that was closest to 3 months.

2.6 Pairwise and Network Meta-analysis

We conducted both pairwise meta-analyses and NMAs using data from the above described sources (i.e., FDA trials and published trials). For pairwise meta-analysis,
we performed a random-effects meta-analysis for all comparisons with two or more trials using the DerSimonian and Laird’s method [18], implemented in STATA package ‘metan’ [19-21]. We assumed both comparison-specific heterogeneity and a common heterogeneity across all comparisons [18]. The latter assumption in theory allows a larger degree of heterogeneity than the comparison-specific heterogeneity approach and would results in less precise estimates.

We used Lu and Ade’s model under a Bayesian hierarchical framework with random-effects for NMA [22, 23]. We conducted the analysis using the R package ‘gemte’ and JAGS for Monte Carlo Markov Chain sampling [24, 25]. The analysis estimated the mean difference in IOP for each pair of interventions. We also estimated the probability for each intervention being ranked at one of the possible positions. Such probabilities were used for plotting the ‘SUCRA’ (surface under the cumulative ranking curve) values [19]. SUCRA value is the cumulative ranking probability, which in theory accounts for the uncertainly in ranking better than a crude ranking probability [19, 26]. Because the numbers of treatments (nodes) in each network were different, the absolute ‘SCURA’ values from the two networks would have different ranges. We therefore used the percentage of ‘SUCRA’, which was defined as ‘SUCRA’ values divided by maximum possible ‘SUCRA’ value in a specific network [26]. A percentage ‘SUCRA’ of 100% means that the treatment has the best overall rankings. After rescaling the ‘SUCRA’ values, we were able to compare the relative rankings of treatments in the two networks regardless of the numbers of treatments in each
network. The larger the SUCRA value the better ranking a drug.

2.7 Evaluation of Clinical and Methodological Heterogeneity

Heterogeneity referred to any variability in the studies included in systematic reviews [3]. We qualitatively evaluate clinical and methodological heterogeneity by comparing the differences in characteristics of the patients, interventions, and the nature of study designs.

2.8 Evaluation of Statistical Heterogeneity and Inconsistency

Statistical heterogeneity is a quantitative measurement of the effect size differences due to variability. It is often measured by $I^2$, $\tau^2$ and Q statistics in conventional pairwise meta-analysis [27]. Q statistics are based on chi$^2$ tests for differences between observed data and expected data. $I^2$ represents the “the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error”[28]. $\tau^2$ is the parameter that captures the between-study variance when using random-effect models.

In the pairwise meta-analysis, we first assumed comparison-specific heterogeneity and estimated the $\tau^2$ for each comparison. Then we assumed that there was a common heterogeneity across all the comparisons, regardless of the interventions compared. Thus, a common $\tau^2$ was estimated to present the heterogeneity of all the studies. In the NMA, we assumed common heterogeneity across the entire network.
and estimated the tau^2.

In the NMA, inconsistency refers to the disagreement in the estimated effect size derived from direct and indirect comparisons. Several methods have been developed to address inconsistency. We used ‘loop-specific’ method and ‘node-splitting’ method to explore inconsistency [29-32]. ‘Loop-specific’ approach checks the triangular or quadratic loops formed by three or four comparators. Inconsistency factors, defined as the difference between estimates derived from the direct and indirect evidence, and their 95% confidence intervals are commonly used to indicate the significance of local inconsistency under the ‘loop-specific’ approach [23, 33]. Another approach, ‘node-splitting’ method, takes out direct comparisons one at a time and conducts a NMA based on the remaining studies to give an indirect estimate for the same comparison. Then the direct and indirect estimates are compared to quantify the amount of inconsistency [29].

2.9 Comparison between Results from Different Sources

We conducted the above described analyses using the FDA trials and trials published in the literature (referred to as published trials) separately. We compared the following results obtained from these two data sources:

1. The point estimates and 95% confidence intervals based on pairwise meta-analyses;

2. The point estimates and 95% credible intervals (intervals where the true parameter
lies in with a probability of 0.95 based on the posterior distribution in Bayesian statistics) based on NMAs;

3. The percentage ‘SUCRA’ values based on NMAs.
3 Results

3.1 Identification of Trials

We identified 26 FDA medical reviews describing results from 72 studies (some are not RCTs). Thirty (42%) RCTs were eligible; however only 16 of 30 (53%) trials reported sufficient data (point estimates and confidence intervals) for our analysis (Figure 1). The network based on FDA trials had 10 nodes (nine active drugs and placebo/vehicle) with a total sample size of 6183 (Figure 2). Timolol was the most commonly used comparator in these trials and was used as the reference group in our analysis.

We identified 105 published trials eligible for analysis from the search results of 10936 studies. The reasons of exclusions are shown in Figure 1. The network based on published trials had 14 nodes (13 active drugs and placebo/vehicle) with a total sample size of 16898 (Figure 2). Similar to the network described above, timolol was the most commonly used comparators (Figure 2).

3.2 Analysis of FDA trials

Table 1 shows the results from direct, pairwise meta-analyses of FDA trials.

Dorzolamide, bimatoprost, unoprostone, levobetaxolol, travoprost were compared to timolol in at least two trials. Timolol lowered IOP more than dorzolamide, unoprostone, and levobetaxolol; the mean differences (95% confidence intervals) in IOP were 1.32 (0.63; 2.00), 2.00 (1.59; 2.41), and 1.25 (0.27; 2.23), respectively.

Bimatoprost and travoprost, the two prostaglandins, on the other hand, lowered IOP
more than timolol; the mean differences (95% confidence intervals) were 2.26 (1.70; 2.82) and 0.98 (0.52; 1.45), respectively.

Table 2 shows the results from the NMA of FDA trials. All active treatments except unoprostone and betaxolol were more effective than placebo/vehicle (i.e., 95% credible interval does not include 0) in lowering IOP at 3 months. Compared to placebo, the mean differences (95% credible intervals) in IOP at 3 months, ordered by the magnitude of IOP reduction, were: bimatoprost 6.03 (4.17; 7.85), travoprost 4.75 (2.95; 6.50), latanoprost 4.60 (2.61; 6.56), timolol 3.76 (2.09; 5.40), brinzolamide 2.66 (0.67; 4.59), levobetaxolol 2.63 (0.75, 3.98), unoprostone 1.77 (-0.06; 3.55), and betaxolol 1.01 (-0.96; 2.95).

Figure 3 shows the probability of each drug being ranked at every possible position and the cumulative ranking probabilities. The ranking results were consistent with the effect estimates obtained from the NMA. Bimatoprost was most likely to be ranked as the best (probability=0.97). Placebo was most likely to be ranked as the worst (probability=0.83). The descending order of ranking, in terms of effectiveness in lowering IOP at 3 months, based on percentage SUCRA value was bimatoprost (9.97), travoprost (8.64), latanoprost (8.31), timolol (7.03), brinzolamide (5.35), dorzolamide (4.70), levobetaxolol (4.57), unoprostone (3.02), betaxolol (2.05) and placebo (0.35).

3.3 Analysis of Published Trials
Table 3 shows the results from direct, pairwise meta-analyses of published trials. Apraclonidine, brimonidine, betaxolol, carteolol, levobunolol, brinzolamide, dorzolamide, bimatoprost, latanoprost, travoprost and unoprostone were compared to timolol in at least two trials. Among these comparisons, six were statistically significant (i.e., 95% confidence interval does not include 0). Timolol lowered IOP more than betaxolol, brinzolamide, dorzolamide; the mean differences (95% confidence intervals) in IOP were 1.58 (0.87; 2.29), 1.10 (0.50; 1.70) and 0.76 (0.13; 1.39), respectively. Bimatoprost, latanoprost and travoprost, the three prostaglandins, on the other hand, lowered IOP more than timolol; the mean differences (95% confidence intervals) were 2.07 (1.49; 2.64), 1.32 (0.88; 1.77) and 1.22 (0.24; 2.20), respectively.

Table 4 shows the results from the NMA of published trials. All active treatments were more effective than placebo/vehicle (i.e., 95% credible interval does not include 0). Compared to placebo, the mean differences (95% credible intervals) in IOP at 3 months, ordered by the magnitude of reduction, were: bimatoprost 5.65 (4.89; 6.40), travoprost 5.02 (4.23; 5.80), latanoprost 4.88 (4.22; 5.54), levobunolol 4.54 (3.77; 5.31), tafluprost 4.41 (2.88; 5.97), timolol 3.70 (3.11; 4.29), brimonidine 3.62 (2.87; 4.37), carteolol 3.44 (2.36; 4.52), dorzolamide 2.56 (1.86; 3.27), apraclonidine 2.55 (0.89; 4.23), brinzomalide 2.46 (1.61; 3.31), unoprostone 1.98 (1.13; 2.84), and betaxolol 2.34 (1.63; 3.05).
Figure 3 shows the probability of each drug being ranked at every possible position and the cumulative ranking probabilities. The ranking results were consistent with the effect estimates gained from the NMA. Bimatoprost was most likely to be ranked as the best (probability=0.924). Placebo was most likely to be ranked as the worst (probability=0.999). The descending order in ranking, in terms of effectiveness in lowering IOP at 3 months, based on percentage SUCRA value was bimatoprost (13.92), travoprost (12.35), latanoprost (11.89), levobunolol (10.82), tafluprost (10.46), timolol (8.42), brimonidine (8.07), carteolol (7.50), dorzolamide (4.79), apraclonidine (4.78), brinzolamide (4.35), betaxolol (3.92), unoprostone (2.74), and placebo (0.00).

3.4 Comparison between FDA Trials and Published Trials

We first compared the results of direct, pairwise meta-analyses based on the two data sources. There were 10 common comparisons between the FDA trials and the published trials. Using results from FDA trials as the reference, we found that, compared to timolol, the results from published trials increased the point estimates of the relative effectiveness of five drugs and decreased the point estimate of the relative effectiveness of one drug. The percentage differences, defined as (estimate of published trials minus estimate of FDA trials) divided by estimate of FDA trials, were 42% for dorzolamide, 8% for unoprostone, 33% for travoprost, 13% for latanoprost, 7% for betaxolol, and -8% for bimatoprost.
We then compared the results derived from NMAs based on the two data sources. There were 9 common treatments between the FDA trials and the published trials, resulting in 36 comparisons. We estimated mean difference in IOP of 36 pairs of these treatments. Figure 5 presents the point estimates from the two data sources. Point estimates from two pairs (light blue dots in Figure 5) showed opposite directions, although the credible intervals all included the null value. Analysis of FDA trials revealed that unoprostone was better than betaxolol by 0.76 (-1.09; 2.57) and brinzolamide was better than dorzolamide by 0.21 (-0.55; 0.97). Analysis of published trials revealed that betaxolol was better than unoprostone by 0.36 (-0.48; 1.20) and dorzolamide was better than brinzolamide by 0.1 (-0.95; 0.74).

The median relative difference of the two networks was 14% (interquartile range: 7% to 42%). The relative differences were greater than 25% in 14 pairs. To account for uncertainty, we also compared the credible intervals (Figure 6). Generally, the results of the published trials had better precision than the results of the FDA trials because of the larger sample size of the published trials (Figure 2). Three comparisons were statistically different based on published trials: unoprostone was better than placebo by 1.98 (1.13; 2.84); latanoprost was better than timolol by 1.18 (0.79; 1.57), and betaxolol was better than placebo by 2.34 (1.63; 3.05). The same three comparisons were not statistically significant based on FDA trials. The relative estimates were 1.77 (-0.06; 3.55) between unoprostone and placebo; 0.84 (-0.23; 1.94) between latanoprost and timolol, and 1.01 (-0.93; 2.95) between betaxolol and placebo.
Figure 7 shows a subset of the 36 comparisons derived from NMAs based on the two data sources using either placebo or timolol as the comparator. Compared to placebo, the point estimates are similar regardless of the data source, although two of the comparisons based on the FDA trials did not reach statistical significance. Compared to timolol, the point estimates are also similar regardless of the data sources, and one comparison based on the FDA trials did not reach statistical significance.

Figure 8 illustrates the differences in percentage ‘SUCRA’ values between two networks. We found that the relative rankings changed for dorzolamide, brinzolamide unoprostone, and betaxolol. The percentage ‘SUCRA’ values, sorted in descending order, for the treatments in the network of FDA trials were: bimatoprost 99.69%, travoprost 86.45%, latanoprost 83.10%, timolol 70.34%, brinzolamide 53.52%, dorzolamide 47.00%, levobetaxolol 45.69%, unoprostone 30.21%, betaxolol 20.48, and placebo 3.54%. The percentage ‘SUCRA’ values, sorted in descending order, for the treatments in the network of published trials were: bimatoprost 99.42%, travoprost 88.18%, latanoprost 84.95%, levobunolol 77.23, tafluprost 74.72%, timolol 60.17%, brimonidine 57.67%, carteolol 53.55%, dorzolamide 34.24%, apraclonidine 34.12%, brinzolamide 31.05%, betaxolol 27.97%, unoprostone 19.56%, and placebo 0.02%.

3.5 Heterogeneity

In general, FDA trials are more homogenous than published trials. The I² values for comparisons based on FDA trials were less than 30%, indicating low level of
statistical heterogeneity [3]. For published trials, we identified two comparisons (betaxolol and placebo, unoprostone and timolol) in which the $I^2$ was greater than 75%, suggesting considerable statistical heterogeneity [3].

3.6 Inconsistency

The network of FDA trials had a “start” shape and did not include comparisons that could be both directly and indirectly estimated. Thus, there was no estimable inconsistency. As for the published trials, using the ‘node-splitting’ approach, we found one node with evidence of statistical inconsistency. Using the ‘loop-specific’ approach, we found 6 out of 34 triangle loops (17.6%) had significant inconsistency (p value < 0.05). We speculated that funding source and extreme effect size could be the reasons introducing inconsistency.
4 Discussions

In this study, we conducted conventional pairwise meta-analyses and NMAs using data from FDA trials and published trials. We found that FDA medical reviews are an important source of clinical trial data that could be used for NMA. The data source does not seem to change the inference about the drug’s relative effectiveness. The relative rankings changed for the middle-ranked drugs but not for the top-ranked drugs.

Several studies have examined the impact of reporting bias using the FDA trials. Turner et al. used the FDA trials and their published versions to describe the selective reporting bias in antidepressants and antipsychotics trials. Based on Turner’s work, Trinquart et al. further examined the reporting bias on NMAs. In both Turner and Trinquart’ studies, they used all FDA trials as one data source and published FDA trials as the second data source, and compared the relative effectiveness and rankings of drugs. However, it is worth noting that non-FDA trials published in the medical literature, the traditional source of trials used for systematic review and meta-analysis, could also be biased. Thus, we framed our question in a broader context than what has been done previously. We used the FDA trials as the ‘reference’ and examined the extent of discrepancies by using all published trials. Because published trials still serve as the main source of data, we answered a pragmatic question: is it worth to put resources in looking for FDA trials when conducting NMA?
Based on our case example, we believe that FDA medical reviews serve as an important source of trial data; systematic reviewers should look for these data for inclusion in their analysis. However, the FDA trial data were incomplete through the public portal. Medical reviews of older drugs are not always available and for one half of the available ones, no precision estimates were provided in the medical reviews. The impact of incomplete FDA trial data on our analysis was threefold. First, the numbers of nodes in the two networks were different. Second, the sample size of some treatments was small in the FDA network (e.g., betaxolol and placebo), resulting in less precise estimates and lower power to detect the differences between the two data sources if there were any. Third, the ‘star’ shape of the FDA network limits the ability to assess inconsistency.

As treatment decisions are increasingly relying on the synthesis of all trials through systematic reviews and meta-analyses, the value of trial data could not be fully realized if there were missing parts, either missing the entire trial or missing information on effect size. We strongly urge the FDA to consider the downstream products of trials and how public is already using the medical reviews to address medical and public health questions. This implies that medical reviews for all approved drugs should be made available. Furthermore, the medical reviews should be prepared following a consistent format and the data on treatment effect should be reported completely. In particular, the precision estimates for the primary and secondary outcomes (or endpoints) should be provided.
We found that the point estimates are similar but the relative rankings are different for a small number of drugs included in our networks. The estimates based on the FDA trials are less precise, which could contribute to the unstable rankings we observed. It is important to interpret the two types of measures (relative effectiveness and rankings) jointly in the context of NMA.

A subset of the FDA trials included in our analyses was also published. One might be interested in comparing the details of the FDA trials with its matching publication to explore disagreement, a traditional approach to examine selective outcome reporting. For the same reason, when we compared the point estimates and relative rankings between the two data sources, we did not account for the precisions because the two data sources are not independent. Future work could focus on developing statistical tests to facilitate such comparisons.

Our study has several strengths. First, this study provided an informative comparison between the results based on published trials and FDA trials in the context of NMA, which was not empirically examined before. We included head-to-head as well as placebo-controlled trials. We answered a pragmatic question of how different the results would be by using different data sources. In addition, NMA allowed the examinations of the impact of data sources and the likelihood of reporting bias between any two treatments in the networks regardless of whether the treatments were
compared directly or indirectly in individual studies.

In conclusion, FDA medical reviews provide a good amount of trial data that should be considered for NMA. The point estimates for the relative effectiveness were similar for most drugs regardless of whether FDA or published trial data are used in the NMA, although the relative rankings changed for some drugs. Whether reporting bias has a role in the differences we observed needs to be further evaluated.
5 References:

14. Li, T., et al., Comparative Effectiveness of First-line Medications for Patients with Primary Open Angle Glaucoma or Ocular Hypertension – A Systematic Review and Network Meta-analysis (in process).
<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>Tianjing Li, P.K.L., MS; Benjamin Rouse; Hwanhee Hong, PhD; Qiyuan Shi; David S. Friedman, MD; Richard Wormald, FRCOphth; Kay Dickersin, PhD, <em>Comparative Effectiveness of First-line Medications for Patients with Primary Open Angle Glaucoma or Ocular Hypertension – A Systematic Review and Network Meta-analysis (in process).</em></td>
</tr>
</tbody>
</table>
6 Appendices

Appendix 1. Search strategies

Cochrane Library
#1 MeSH descriptor: [Glaucoma, Open-Angle] explode all trees
#2 MeSH descriptor: [Ocular Hypertension] explode all trees
#3 (open near/2 angle near/2 glaucoma*)
#4 (POAG or OHT)
#5 ((increas* or elevat* or high*) near/3 (ocular or intra-ocular)) and pressure
#6 [34-#5]
#7 MeSH descriptor: [Adrenergic beta-Antagonists] explode all trees
#8 MeSH descriptor: [Timolol] explode all trees
#9 Timolol*
#10 MeSH descriptor: [Metipranolol] explode all trees
#11 Metipranolol*
#12 MeSH descriptor: [Carteolol] explode all trees
#13 Carteolol*
#14 MeSH descriptor: [Levobunolol] explode all trees
#15 Levobunolol*
#16 MeSH descriptor: [Betaxolol] explode all trees
#17 Betaxolol*
#18 MeSH descriptor: [Carbonic Anhydrase Inhibitors] explode all trees
#19 (Carbonic near/2 Anhydrase near/2 Inhibitor*)
#20 MeSH descriptor: [Acetazolamide] explode all trees
#21 Acetazolam*
#22 Brinzolamide*
#23 Dorzolamide*
#24 MeSH descriptor: [Prostaglandins, Synthetic] explode all trees
#25 latanoprost*
#26 travoprost*
#27 bimatoprost*
#28 unoprostone*
#29 tafluprost*
#30 MeSH descriptor: [Antihypertensive Agents] explode all trees
#31 MeSH descriptor: [Pilocarpine] explode all trees
#32 Pilocarpin*
#33 MeSH descriptor: [Epinephrine] explode all trees
#34 epinephrine*
#35 dipivefrin*
#36 MeSH descriptor: [Adrenergic alpha-2 Receptor Agonists] explode all trees
#37 (adrenergic near/2 alpha* near/3 agonist*)
#38 apraclonidin*
26

#39 brimonidine*
#40 (drug* or medic* or pharmacologic*) near/3 (treat* or therap* or intervent*)
#41 {or #7-#40}
#42 #6 and #41

MEDLINE (OVID)
1. exp clinical trial/ [publication type]
2. (randomized or randomised).ab,ti.
3. placebo.ab,ti.
4. dt.fs.
5. randomly.ab,ti.
6. trial.ab,ti.
7. groups.ab,ti.
8. or/1-7
9. exp animals/
10. exp humans/
11. 9 not (9 and 10)
12. 8 not 11
13. exp glaucoma open angle/
14. exp ocular hypertension/
15. (open adj2 angle adj2 glaucoma$).tw.
16. (POAG or OHT).tw.
17. ((increas$ or elevat$ or high$) adj3 (ocular or intra-ocular)) and pressure).tw.
18. or/13-17
19. exp adrenergic beta antagonists/
20. exp timolol/
21. timolol$.tw.
22. exp metipranolol/
23. metipranolol$.tw.
24. exp carteolol/
25. carteolol$.tw.
26. exp levobunolol/
27. levobunolol$.tw.
28. exp betaxolol/
29. betaxolol$.tw.
30. exp carbonic anhydrase inhibitors/
31. (carbonic adj2 anhydrase adj2 inhibitor$).tw.
32. exp Acetazolamide/
33. acetazolamide$.tw.
34. brinzolamide$.tw.
35. dorzolamide$.tw.
36. exp Prostaglandins, Synthetic/
37. latanoprost$.tw.
38. travoprost$.tw.
39. bimatoprost$.tw.
40. unoprostone$.tw.
41. brimonidine$.tw.
42. exp antihypertensive agents/
43. exp pilocarpine/
44. pilocarpin$.tw.
45. exp epinephrine/
46. epinephrin$.tw.
47. dipivefrin$.tw.
48. exp Adrenergic alpha-2 Receptor Agonists/
49. ((adrenergic adj2 alpha$ adj2 receptor$) or (adrenergic adj2 alpha$ adj2 agonist$)).tw.
50. apraclonidin$.tw.
51. tafluprost$.tw.
52. ((drug$ or medic$ or pharmacologic$) adj3 (treat$ or therap$ or intervent$)).tw.
53. or/19-52
54. 18 and 53
55. 12 and 54

**Embase.com**
#1 'randomized controlled trial'/exp
#2 'randomization'/exp
#3 'double blind procedure'/exp
#4 'single blind procedure'/exp
#5 random*:ab,ti
#6 #1 OR #2 OR #3 OR #4 OR #5
#7 'animal'/exp OR 'animal experiment'/exp
#8 'human'/exp
#9 #7 AND #8
#10 #7 NOT #9
#11 #6 NOT #10
#12 'clinical trial'/exp
#13 (clin* NEAR/3 trial*):ab,ti
#14 ((singl* OR doubl* OR trebl* OR tripl*) NEAR/3 (blind* OR mask*)):ab,ti
#15 'placebo'/exp
#16 placebo*:ab,ti
#17 random*:ab,ti
#18 'experimental design'/exp
#19 'crossover procedure'/exp
#20 'control group'/exp
#21 'latin square design'/exp
#22 #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21
#23 #22 NOT #10
#24 #23 NOT #11
#25 'comparative study'/exp
#26 'evaluation'/exp
#27 'prospective study'/exp
#28 control*:ab,ti OR prospectiv*:ab,ti OR volunteer*:ab,ti
#29 #25 OR #26 OR #27 OR #28
#30 #29 NOT #10
#31 #30 NOT (#11 OR #23)
#32 #11 OR #24 OR #31
#33 'open angle glaucoma'/exp
#34 'intraocular hypertension'/exp
#35 (open NEAR/2 angle):ab,ti AND (angle NEAR/2 glaucoma*):ab,ti
#36 poag:ab,ti OR oht:ab,ti
#37 ((increas* OR elevat* OR high*) NEAR/3 (ocular OR 'intra ocular')):ab,ti AND pressure:ab,ti
#38 #33 OR #34 OR #35 OR #36 OR #37
#39 'beta adrenergic receptor blocking agent'/exp
#40 'timolol'/exp
#41 timolol*:ab,ti
#42 'metipranolol'/exp
#43 metipranolol*:ab,ti
#44 'carteolol'/exp
#45 carteolol*:ab,ti
#46 'levobunolol'/exp
#47 levobunolol*:ab,ti
#48 'betaxolol'/exp
#49 betaxolol*:ab,ti
#50 'carbonate dehydratase inhibitor'/exp
#51 (carbonic NEAR/2 anhydrase):ab,ti AND (anhydrase NEAR/2 inhibitor*):ab,ti
#52 'acetazolamide'/exp
#53 acetazolamide*:ab,ti
#54 brinzolamide*:ab,ti
#55 dorzolamide*:ab,ti
#56 'latanoprost'/exp
#57 latanoprost*:ab,ti
#58 'travoprost'/exp
#59 travoprost*:ab,ti
#60 'bimatoprost'/exp
#61 bimatoprost*:ab,ti
#62 'unoprostone isopropyl ester'/exp
#63 unoprostone*:ab,ti
#64 'brimonidine'/exp
#65 brimonidine*:ab,ti
#66 'antihypertensive agent'/exp
#67 'pilocarpine'/exp
#68 pilocarpin*:ab,ti
 PubMed

#1 ((randomized controlled trial[pt]) OR (controlled clinical trial[pt]) OR (randomised[tiab] OR randomized[tiab]) OR (placebo[tiab]) OR (drug therapy[sh]) OR (randomly[tiab]) OR (trial[tiab]) OR (groups[tiab])) NOT (animals[mh] NOT humans[mh])
#2 (open[tw] AND angle[tw] AND glaucoma*[tw]) NOT Medline[sb]
#3 (POAG[tw] OR OHT[tw]) NOT Medline[sb]
#4 (((increase*[tw] OR elevat*[tw] OR high*[tw]) AND (ocular[tw] OR intra-ocular[tw])) AND pressure[tw]) NOT Medline[sb]
#5 #2 OR #3 OR #4
#6 timolol*[tw] NOT Medline[sb]
#7 metipranolol*[tw] NOT Medline[sb]
#8 carteolol*[tw] NOT Medline[sb]
#9 levbunolol*[tw] NOT Medline[sb]
#10 betaxolol*[tw] NOT Medline[sb]
#11 (carbonic[tw] AND anhydrase[tw] AND inhibitor*[tw]) NOT Medline[sb]
#12 acetazolamide*[tw] NOT Medline[sb]
#13 brinzolamide*[tw] NOT Medline[sb]
#14 dorzolamide*[tw] NOT Medline[sb]
#15 latanoprost*[tw] NOT Medline[sb]
#16 travoprost*[tw] NOT Medline[sb]
#17 bimatoprost*[tw] NOT Medline[sb]
#18 unoprostone*[tw] NOT Medline[sb]
#19 brimonidine*[tw] NOT Medline[sb]
#20 pilocarpin*[tw] NOT Medline[sb]
#21 epinephrin*[tw] NOT Medline[sb]
#22 dipivefrin* NOT Medline[sb]
#24 apraclonidin*[tw] NOT Medline[sb]
#25 tafluprost*[tw] NOT Medline[sb]
#26 ((drug*[tw] OR medic*[tw] OR pharmacologic*[tw]) AND (treat*[tw] OR therap*[tw] OR intervent*[tw])) NOT Medline[sb]
#27 #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26
#28 #5 AND #27
#29 #1 AND #28

**FDA (Drugs@FDA)**

#1. Acetazolamide
#2. Brimonidine
#3. Betaxolol
#4. Dichlorphenamide
#5. Levobunolol
#6. Timolol
#7. Brinzolamide
#8. Dorzolamide
#9. Bimatoprost
#10. Latanoprost
#11. Travoprost
#12. Unoprostone
#13. Acetazolamide
#14. Methazolamide
#15. Pilocarpine
#16. Carbachol
#17. Echothiophate
#18. Epinephrine
#19. Dipivefrin
#20. Metipranolol
#21. Demecarium
#22. Apraclonidine
#23. Carteolol
#24. Tafluprost
7 Tables

Table 1. Summary estimates for intraocular pressure at 3 months derived from pair-wise meta-analysis based on 13 direct comparisons from 16 FDA trials

<table>
<thead>
<tr>
<th>Comparison-specific heterogeneity</th>
<th>Common heterogeneity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. of trials</td>
<td>Mean difference</td>
</tr>
<tr>
<td>Placebo vs. Levobetaxolol</td>
<td>1</td>
</tr>
<tr>
<td>Levobetaxolol</td>
<td>1</td>
</tr>
<tr>
<td>Timolol</td>
<td>1</td>
</tr>
<tr>
<td>Timolol vs. Dorzolamide</td>
<td>3</td>
</tr>
<tr>
<td>Timolol vs. Bimatoprost</td>
<td>2</td>
</tr>
<tr>
<td>Timolol vs. Unoprostone</td>
<td>2</td>
</tr>
<tr>
<td>Timolol vs. Levobetaxolol</td>
<td>3</td>
</tr>
<tr>
<td>Timolol vs. Travoprost</td>
<td>3</td>
</tr>
<tr>
<td>Timolol vs. Latanoprost</td>
<td>1</td>
</tr>
<tr>
<td>Betaxolol vs. Levobetaxolol</td>
<td>1</td>
</tr>
<tr>
<td>Betaxolol vs. Timolol</td>
<td>1</td>
</tr>
<tr>
<td>Brinzolamide vs. Dorzolamide</td>
<td>2</td>
</tr>
<tr>
<td>Latanoprost vs. Travoprost</td>
<td>1</td>
</tr>
</tbody>
</table>

*Estimated tau square is 0.039; estimated I square is 36.03%

Tau square: between-study variance in random-effect models; I square: proportion of variance due to heterogeneity
Table 2. Summary estimates for intraocular pressure at 3 months derived from network meta-analysis of 16 FDA trials

<table>
<thead>
<tr>
<th>Placebo</th>
<th>Betaxolol</th>
<th>(-0.93;2.95)</th>
<th>3.76</th>
<th>2.75</th>
<th>Timolol</th>
<th>(2.09;5.40)</th>
<th>1.35</th>
<th>-1.40</th>
<th>Levobetaxolol</th>
<th>(-0.3;3.03)</th>
<th>(0.75;3.98)</th>
<th>2.66</th>
<th>1.65</th>
<th>-1.10</th>
<th>0.30</th>
<th>Brinzolamide</th>
<th>(-2.06;0.62)</th>
<th>-1.40</th>
<th>-0.62</th>
<th>0.30</th>
<th>-0.62</th>
<th>0.30</th>
<th>-0.62</th>
<th>0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>Betaxolol</td>
<td>(-0.93;2.95)</td>
<td>3.76</td>
<td>2.75</td>
<td>Timolol</td>
<td>(2.09;5.40)</td>
<td>1.35</td>
<td>-1.40</td>
<td>Levobetaxolol</td>
<td>(-0.3;3.03)</td>
<td>(0.75;3.98)</td>
<td>2.66</td>
<td>1.65</td>
<td>-1.10</td>
<td>0.30</td>
<td>Brinzolamide</td>
<td>(-2.06;0.62)</td>
<td>-1.40</td>
<td>-0.62</td>
<td>0.30</td>
<td>-0.62</td>
<td>0.30</td>
<td>-0.62</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Grey

Green

Red

Blue

Unit: mmHg; Color coding: drug class
Mean difference < 0 favors the drug in the column
Mean difference > 0 favors the drug in the row
Reported numbers are calculated by column - row under the Lu and Ades homogeneous random effects model assuming consistency
Reported posterior means and 95% Bayesian credible intervals

Placebo/vehicle/no treatment
Beta-blocker
Carbonic anhydrase inhibitor
Prostaglandin analog

32
Table 3. Summary estimates for intraocular pressure at 3 months derived from pair-wise meta-analysis based on 36 direct comparisons from 105 published trials

<table>
<thead>
<tr>
<th>Comparison-specific heterogeneity</th>
<th>Common heterogeneity*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Num. of trials</td>
</tr>
<tr>
<td>Placebo vs.</td>
<td></td>
</tr>
<tr>
<td>Brimonidine</td>
<td>1</td>
</tr>
<tr>
<td>Betaxolol</td>
<td>2</td>
</tr>
<tr>
<td>Levobunolol</td>
<td>2</td>
</tr>
<tr>
<td>Timolol</td>
<td>4</td>
</tr>
<tr>
<td>Brinzolamide</td>
<td>2</td>
</tr>
<tr>
<td>Dorzolamide</td>
<td>4</td>
</tr>
<tr>
<td>Bimatoprost</td>
<td>1</td>
</tr>
<tr>
<td>Unoprostone</td>
<td>1</td>
</tr>
<tr>
<td>Apraclonidine vs.</td>
<td></td>
</tr>
<tr>
<td>Timolol</td>
<td>2</td>
</tr>
<tr>
<td>Betaxolol</td>
<td>1</td>
</tr>
<tr>
<td>Timolol</td>
<td>4</td>
</tr>
<tr>
<td>Brinzolamide</td>
<td>2</td>
</tr>
<tr>
<td>Latanoprost</td>
<td>5</td>
</tr>
<tr>
<td>Travoprost</td>
<td>1</td>
</tr>
<tr>
<td>Betaxolol vs.</td>
<td></td>
</tr>
<tr>
<td>Levobunolol</td>
<td>2</td>
</tr>
<tr>
<td>Timolol</td>
<td>8</td>
</tr>
<tr>
<td>Dorzolamide</td>
<td>2</td>
</tr>
<tr>
<td>Latanoprost</td>
<td>2</td>
</tr>
<tr>
<td>Unoprostone</td>
<td>1</td>
</tr>
<tr>
<td>Carteolol vs.</td>
<td></td>
</tr>
<tr>
<td>Levobunolol</td>
<td>1</td>
</tr>
<tr>
<td>Timolol</td>
<td>4</td>
</tr>
<tr>
<td>Levobunolol vs.</td>
<td></td>
</tr>
<tr>
<td>Timolol</td>
<td>11</td>
</tr>
<tr>
<td>Timolol vs.</td>
<td></td>
</tr>
<tr>
<td>Brinzolamide</td>
<td>3</td>
</tr>
<tr>
<td>Dorzolamide</td>
<td>5</td>
</tr>
<tr>
<td>Bimatoprost</td>
<td>5</td>
</tr>
<tr>
<td>Latanoprost</td>
<td>14</td>
</tr>
<tr>
<td>Travoprost</td>
<td>5</td>
</tr>
<tr>
<td>Unoprostone</td>
<td>1</td>
</tr>
<tr>
<td>Brinzolamide vs.</td>
<td></td>
</tr>
<tr>
<td>Dorzolamide</td>
<td>2</td>
</tr>
<tr>
<td>Latanoprost</td>
<td>1</td>
</tr>
<tr>
<td>Bimatoprost vs.</td>
<td></td>
</tr>
<tr>
<td>Latanoprost</td>
<td>6</td>
</tr>
<tr>
<td>Travoprost</td>
<td>8</td>
</tr>
<tr>
<td>Latanoprost vs.</td>
<td></td>
</tr>
<tr>
<td>Travoprost</td>
<td>7</td>
</tr>
<tr>
<td>Tafluprost</td>
<td>1</td>
</tr>
<tr>
<td>Unoprostone</td>
<td>6</td>
</tr>
</tbody>
</table>

*Estimated tau squared is 0.4374; estimated I-squared is 59.63%*  
**Tau square: between-study variance in random-effect models; I squared: proportion of variance due to heterogeneity**
Table 4. Summary estimates for intraocular pressure at 3 months derived from network meta-analysis of 105 published trials

<table>
<thead>
<tr>
<th>Drug</th>
<th>Mean difference</th>
<th>95% Bayesian credible intervals</th>
<th>Mean difference</th>
<th>95% Bayesian credible intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>2.55</td>
<td>(0.89;4.23)</td>
<td>3.44</td>
<td>(2.36;4.52)</td>
</tr>
<tr>
<td>Apraclonidine</td>
<td>1.07</td>
<td>(0.53;2.77)</td>
<td>-0.21</td>
<td>(-1.03;0.60)</td>
</tr>
<tr>
<td>Dibonidine</td>
<td>1.28</td>
<td>(0.47;3.07)</td>
<td>1.28</td>
<td>(0.50;3.05)</td>
</tr>
<tr>
<td>Betaxolol</td>
<td>1.10</td>
<td>(0.58;2.62)</td>
<td>0.19</td>
<td>(0.00;0.44)</td>
</tr>
<tr>
<td>Carteolol</td>
<td>1.09</td>
<td>(0.59;2.59)</td>
<td>2.19</td>
<td>(0.67;3.71)</td>
</tr>
<tr>
<td>Levocabumol</td>
<td>0.27</td>
<td>(-0.84;1.38)</td>
<td>1.36</td>
<td>(0.86;2.87)</td>
</tr>
<tr>
<td>Timolol</td>
<td>-0.84</td>
<td>(-2.60;0.92)</td>
<td>-0.27</td>
<td>(-2.13;0.61)</td>
</tr>
<tr>
<td>Brinzolamide</td>
<td>1.10</td>
<td>(0.50;2.59)</td>
<td>0.08</td>
<td>(0.00;0.16)</td>
</tr>
<tr>
<td>Dorzolamide</td>
<td>1.14</td>
<td>(0.50;2.77)</td>
<td>1.98</td>
<td>(1.10;3.91)</td>
</tr>
<tr>
<td>Latanoprost</td>
<td>3.16</td>
<td>(2.08;4.24)</td>
<td>2.32</td>
<td>(1.31;3.33)</td>
</tr>
<tr>
<td>Travoprost</td>
<td>3.08</td>
<td>(2.10;4.06)</td>
<td>1.26</td>
<td>(0.35;2.17)</td>
</tr>
<tr>
<td>Tafluprost</td>
<td>2.32</td>
<td>(1.43;3.20)</td>
<td>1.13</td>
<td>(0.22;2.02)</td>
</tr>
<tr>
<td>Unoprostone</td>
<td>2.15</td>
<td>(1.25;3.04)</td>
<td>0.13</td>
<td>(0.00;0.26)</td>
</tr>
<tr>
<td>Bimatoprost</td>
<td>2.10</td>
<td>(1.20;3.00)</td>
<td>1.13</td>
<td>(0.22;2.02)</td>
</tr>
<tr>
<td>Latanoprost</td>
<td>5.02</td>
<td>(4.04;6.01)</td>
<td>0.13</td>
<td>(0.00;0.26)</td>
</tr>
<tr>
<td>Travoprost</td>
<td>4.41</td>
<td>(3.51;5.31)</td>
<td>1.13</td>
<td>(0.22;2.02)</td>
</tr>
<tr>
<td>Tafluprost</td>
<td>3.08</td>
<td>(2.10;4.06)</td>
<td>1.13</td>
<td>(0.22;2.02)</td>
</tr>
<tr>
<td>Unoprostone</td>
<td>3.16</td>
<td>(2.08;4.24)</td>
<td>1.26</td>
<td>(0.35;2.17)</td>
</tr>
<tr>
<td>Placebo/vehicle/no treatment</td>
<td>2.55</td>
<td>(0.89;4.23)</td>
<td>3.44</td>
<td>(2.36;4.52)</td>
</tr>
<tr>
<td>Alpha-2 adrenergic agonist</td>
<td>1.07</td>
<td>(0.53;2.77)</td>
<td>-0.21</td>
<td>(-1.03;0.60)</td>
</tr>
<tr>
<td>Beta-blocker</td>
<td>1.28</td>
<td>(0.47;3.07)</td>
<td>1.28</td>
<td>(0.50;3.05)</td>
</tr>
<tr>
<td>Carbonic anhydrase inhibitor</td>
<td>1.10</td>
<td>(0.58;2.62)</td>
<td>0.19</td>
<td>(0.00;0.44)</td>
</tr>
<tr>
<td>Prostaglandin analog</td>
<td>1.09</td>
<td>(0.59;2.59)</td>
<td>2.19</td>
<td>(0.67;3.71)</td>
</tr>
</tbody>
</table>

Mean difference < 0 favors the drug in the column
Mean difference > 0 favors the drug in the row
Reported numbers are calculated by column - row under the Lu and Ades homogeneous random effects model assuming consistency
Reported posterior means and 95% Bayesian credible intervals

Unit: mmHg; Color coding: drug class
Grey: Placebo/vehicle/no treatment
Gold: Alpha-2 adrenergic agonist
Green: Beta-blocker
Red: Carbonic anhydrase inhibitor
Blue: Prostaglandin analog
Table 5. Comparisons between the results of direct pairwise meta-analyses based on FDA trials and published trials

<table>
<thead>
<tr>
<th></th>
<th>Pairwise comparisons based on FDA trials</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Effect size (95%CI)</td>
<td>Tau-squared</td>
<td>I-squared</td>
<td>N</td>
</tr>
<tr>
<td>Placebo vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betaxolol</td>
<td>1</td>
<td>-1.00 (-2.76; 0.76)</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Timolol</td>
<td>1</td>
<td>-2.70 (-4.44;-0.96)</td>
<td>NA</td>
<td>NA</td>
<td>4</td>
</tr>
<tr>
<td>Timolol vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dorzolamide</td>
<td>3</td>
<td>1.32 (0.63; 2.00)</td>
<td>0.12</td>
<td>24%</td>
<td>5</td>
</tr>
<tr>
<td>Bimatoprost</td>
<td>2</td>
<td>-2.26 (-2.82;-1.70)</td>
<td>0.00</td>
<td>0%</td>
<td>5</td>
</tr>
<tr>
<td>Unoprostone</td>
<td>2</td>
<td>2.00 (1.59;2.41)</td>
<td>0.00</td>
<td>0%</td>
<td>3</td>
</tr>
<tr>
<td>Travoprost</td>
<td>3</td>
<td>-0.98 (-1.45;-0.52)</td>
<td>0.03</td>
<td>19%</td>
<td>5</td>
</tr>
<tr>
<td>Latanoprost</td>
<td>1</td>
<td>-1.10 (-1.99;-0.21)</td>
<td>NA</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>Betaxolol</td>
<td>1</td>
<td>1.70 (0.05;-3.45)</td>
<td>NA</td>
<td>NA</td>
<td>8</td>
</tr>
<tr>
<td>Brinzolamide vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dorzolamide</td>
<td>2</td>
<td>0.22 (-0.37;0.80)</td>
<td>0.06</td>
<td>32%</td>
<td>2</td>
</tr>
<tr>
<td>Latanoprost vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travoprost</td>
<td>1</td>
<td>-0.30 (-1.03;0.43)</td>
<td>NA</td>
<td>NA</td>
<td>7</td>
</tr>
</tbody>
</table>
8 Figures

Figure 1. Identification of trials

POAG: primary open angle glaucoma
FDA: Food and Drug Administration
RCT: randomized clinical trial
Figure 2. Network graphs

FDA trials:
16 RCTs
6183 Participants
10 Treatments

Published trials:
105 RCTs
16898 Participants
14 Treatments

Colors indicate the drug classes as follows: Green: Beta-blocker; Orange Red: Carbonic anhydras inhibitor; Orange: Alpha-2 adrenergic agonist; Blue: Prostaglandin analog; Grey: Placebo/vehicle/no treatment. The size of nodes reflects the number of participants in each treatment group and the thickness of edges is proportional to the number of trials.
Figure 3. Ranking probabilities for any drug at any position in the FDA trials network
Figure 4. Ranking probabilities for any drug at any position in the published trials network
Figure 5. Scatterplot of relative effect size between two treatments based on both FDA and published trials

The relative effect size is measured as the differences in IOP in mmHg at 3 months between two treatments. The solid line is 45° line. The dashed lines are ±25% lines. The blue points are the relative effect sizes that are agreeing directions based on two data sources. They are unoprostone vs betaxolol and dorzolamide vs brinzolamide.
Figure 6. Relative effect size based on FDA trials and published trials

![Graph showing differences in IOP at 3 months for various comparisons between medications. The x-axis represents differences in IOP in mmHg, ranging from -8 to 5. The y-axis lists medication comparisons. The graph includes data from FDA trials and published trials, indicated by different markers and colors.]
Figure 7. Relative effect size between active drugs and placebo/timolol based on FDA trials and published trials
Figure 8. Percentage ‘SUCRA’ of drugs in the two networks
Curriculum vitae

PERSONAL
Qiyuan Shi was born on Oct. 21st 1990 in Shanghai, China.

EDUCATION
Johns Hopkins University School of Public Health
09/2013-05/2015
Degree: Master of Health Science
Concentration: Epidemiology, Clinical trials and Evidence Synthesis

Fudan University School of Pharmacy
09/2009-06/2013
Major: Pharmacy
Degree: Bachelor of Science

PROFESSIONAL DEVELOPMENT
Software & Languages: Microsoft Office Suite, SAS (Certified Base Programmer), R, STATA, ArcGIS
Typesetting: Latex
Language: English & Mandarin

PUBLIC HEALTH EXPERIENCE
Research Assistant, Cochrane Eyes and Vision Group Johns Hopkins University Center for Clinical Trials, Department of Epidemiology, School of Public Health, Johns Hopkins University
02/2014—present
- Search, collect, manage and screen clinical research studies from different databases
- Extract and manage data from studies
- Conduct qualitative and quantitative analysis

Research Assistant, Cochrane Hypertension Group Fudan University
Department of Clinical Pharmacology, School of Pharmacy, Fudan University
06/2012—06/2013
- Search, collect, manage and screen clinical research studies from different databases
- Extract and manage data from studies
- Conduct qualitative and quantitative analysis

TEACHING EXPERIENCE
Teaching assistant and lab instructor, Department of Epidemiology & Department of Biostatistics, Johns Hopkins University

09/2014—12/2014

INTERNSHIP EXPERIENCE
Pharmacist
Zhongshan Hospital, Shanghai, China
07/2012-12/2012