Methods and results

Follow the procedures in PRISMA 2009.

10. Data extraction

An undergraduate assistant manually extracted data from the tables of the identified studies. In three cases where the data of interest was not cited in - the IOP measurements and medication numbers in (ref Liu et al.), and the medication numbers in (ref Hayashi, Tham 2013), the data was extracted from figures using a graphical data extraction tool (WebPlotDigitizer, http://arohatgi.info/WebPlotDigitizer/; ref Cochrane https://www.ncbi.nlm.nih.gov/pubmed/26780258).

The data tables were spot-checked by M.M. against the primary studies. Data was also audited for self-consistency by P.M. – e.g. ensuring that differences in IOP post-surgery were consistent with pre- and post-surgery IOP measurements, that number of eyes tracked did not increase through follow-up, etc. The full extent of these automated checks can be viewed at (ref github link).

The data for OAG studies was also cross-verified against Table 1 of Thomas et al (ref). The full extracted data can be accessed at (ref Github link).

11 & 13. Data items & summary measures

We extracted the following columns from the primary sources:

```
extracted.data <- read.column.names()
kable(data.frame(extracted.data))
```

extracted.data

Journal, Volume (Page)
Author
Study Type
Year
MIGs (Y or N)
Wash Out
Wash out baseline
Baseline wash out SD
Wash out IOP period
Wash out IOP
Wash out SD
Types of MIGS (if any)
Age (Mean)
Age (Std Dev)
% Male
% Female
% OAG
% ACG
% NTG
% PXG
acuteangleclosure
Note on Type of Glaucoma
PreOpEyes
PreOpIOPMean
PreOpIOPStdDev
Eyes6mo
MeanIOP6mo
MeanIOPsd6mo

IOPchangemean6m
IOPchangesd6mo
Eyes12mo
MeanIOP12mo
MeanIOPsd12mo
IOPchangemean12m
IOPchangeSD12mo
LastPeriodofEyes
TimeofLastPostOp
LastPeriodIOPMean
LastPeriodIOPStdDev
LastPeriodAbsIOPChangeMean LastPeriodAbsIOPChangeStd
RxPreOpMean
RxPreOpStdDev
RxPostOpMean
RxPostOpStdDev
VisualAcuityPreOpMean
VisualAcuityPreOpStdDev
VisualAcuityPostOpMean
VisualAcuityPostOpStdDev
Notes
Exclude
Exclusion notes

TODO(Marisse): transform this to a table, group similar measurements together.

The main metric of interest for our meta-analysis was the IOP change at the final follow-up period (12 months or longer). In complementary analyses, we performed meta-analyses separately for 6 month, 12 month, and > 12 month follow-ups. When the final period of follow up was not equal for every patient, we put in the 6-month follow up bin if it was less than 12 months, on average, and in the > 12 month bin it was more than 12 months.

We also analyzed the change in number of medications between the pre- period and final follow-up.

Aggregation and cleaning

Some studies used washout in at least some time periods to measure the effect of surgery separately from the effect of medication. When both pre- and post-surgery IOP were reported with washout, we used the washed out IOPs (refs Pfeiffer, Vold); this is indicated by an asterisk in the main results. When washout was used in only some periods, we used non-washout values across time periods, for consistency. The only exception was (ref Azuara−Blanco 2016), where IOP was measured with washout in the pre-period but not in the post period, due to study design. We indicate this study with a double asterisk.

Many studies did not report IOP change directly, but rather only absolute IOP in the pre-period and at the follow-up period. Assuming follow-ups are missing completely at random (MCAR), the IOP change can be estimated simply by the difference of these measurements, while the SD (σ) of the IOP change is estimated by:

$\sigma^2_{drop} = \sigma^2_{pre} + \sigma^2_{post} - 2r\sigma_{pre}\sigma_{post}$

r, the correlation between the pre-and-post measurements, was set to the median of the correlations in the studies reporting the full set of metrics (r = 0.28).

```
df <- read.data(impute.change = FALSE) %>% filter(subtype != 'acute')
rs <- with(df, (PreOpIOPStdDev ** 2 + LastPeriodIOPStdDev ** 2 - LastPeriodAbsIOPChangeStdDev ** 2) / (PreOpIO
quantile(rs, .5, na.rm=TRUE)
```

```
##       50%
## 0.2788965
```

The reported SD of the change in IOP in one study (Jacobi et al. 2002) implied a pre-post correlation larger than 1; since the reported range of the IOP change was also incompatible with the reported range, we replaced this SD with one estimated from the reported pre- and post-surgery IOP SDs.

Many studies reported several arms corresponding to different severities, treatments and subtypes. For phacoemulsification studies, we aggregated the data to obtain one arm per glaucoma subtype (ACG, OAG, PXG, acute) per study.

After this step, the total number of arms and eyes per subtype was as follows:

```
df <- read.data(fill.last=FALSE)
df_ <- df %>% filter(!is.na(LastPeriodAbsIOPChangeMean)) %>% group_by(subtype) %>%
  dplyr::summarize(
    n.total = n(),
    n.prospective = sum(StudyType == 'Prospective'),
    n.retrospective = sum(StudyType == 'Retrospective'),
    PreOpEyes = sum(PreOpEyes),
    FinalPeriodEyes = sum(LastPeriodEyes))

df_ <- df_ %>% rbind(., summarize(df_,
                                  subtype="(all)",
                                  n.total=sum(n.total),
                                  n.prospective=sum(n.prospective),
                                  n.retrospective=sum(n.retrospective),
                                  PreOpEyes=sum(PreOpEyes),
                                  FinalPeriodEyes = sum(FinalPeriodEyes)))
kable(df_, caption = "Number of arms with 12 month+ follow up")
```

Table 1: Number of arms with 12 month+ follow up

subtype	n.total	n.prospective	n.retrospective	PreOpEyes	FinalPeriodEyes
ACG	9	6	3	522	462
acute	3	2	1	76	74
OAG	14	6	8	794	686
PXG	4	2	2	125	114
(all)	30	16	14	1517	1336

```
df_ <- df %>% group_by(subtype) %>%
  dplyr::summarize(
    n.total = n(),
    n.prospective = sum(StudyType == 'Prospective'),
    n.retrospective = sum(StudyType == 'Retrospective'),
    PreOpEyes = sum(PreOpEyes))

df_ <- df_ %>% rbind(., summarize(df_,
                                  subtype="(all)",
                                  n.total=sum(n.total),
                                  n.prospective=sum(n.prospective),
                                  n.retrospective=sum(n.retrospective),
                                  PreOpEyes=sum(PreOpEyes)))
kable(df_, caption = "Number of arms across all studies")
```

Table 2: Number of arms across all studies

subtype	n.total	n.prospective	n.retrospective	PreOpEyes
ACG	18	12	6	898
acute	5	4	1	144
OAG	18	8	10	940
PXG	4	2	2	125
(all)	45	26	19	2107

The full list of studies is as follow:

```
df_ <- df %>% filter(!is.na(LastPeriodAbsIOPChangeMean)) %>%
  arrange(subtype, Year, study.name) %>%
  dplyr::select(
```

```
    study.name,
    Year,
    StudyType,
    subtype,
    PreEyes=PreOpEyes,
    PreIOPMean=PreOpIOPMean,
    PreIOPSD=PreOpIOPStdDev,
    TimeofLastPostOp)
kable(df_, digits = 1)
```

study.name	Year	StudyType	subtype	PreEyes	PreIOPMean	PreIOPSD	TimeofLastPostOp
Hayashi et al. (2001)†	2001	Prospective	ACG	74	21.4	3.9	24 mo
Mierzejewski et al. (2008)	2008	Retrospective	ACG	25	19.5	5.2	14.5 +/- 4.4 mo
Tham et al. (2008)	2008	Prospective	ACG	35	16.3	3.0	24 mo
Tham et al. (2009)	2009	Prospective	ACG	27	24.4	6.1	24 mo
Liu et al. (2011)	2011	Retrospective	ACG	56	16.4	4.0	24 mo
Tham et al. (2013)	2013	Prospective	ACG	26	24.1	4.1	24 mo
Dias-Santos (2015)	2015	Prospective	ACG	15	19.9	8.3	31.1 +/-4.9 mo
Azuara-Blanco et al. (2016)**	2016	Prospective	ACG	208	29.5	8.2	36 mo
Lee et al. (2016)	2016	Retrospective	ACG	56	14.4	4.6	43.45 +/- 17.5 mo
Lam et al. (2008)	2008	Prospective	acute	31	59.7	8.7	18 mo
Lee et al. (2010)	2010	Retrospective	acute	26	49.0	10.4	48 mo
Husain et al. (2012)	2012	Prospective	acute	19	57.4	16.9	24 mo
Kim et al. (1999)	1999	Retrospective	OAG	31	18.1	3.1	16.4 mo
Hayashi et al. (2001)†	2001	Prospective	OAG	68	20.7	5.4	24 mo
Leelachaikul et al. (2005)	2005	Retrospective	OAG	58	16.5	3.8	18 mo
Mathalone et al. (2005)	2005	Retrospective	OAG	58	17.0	4.6	24 mo
Damji et al. (2006)	2006	Prospective	OAG	29	18.5	3.5	24 mo
Shingleton et al. (2006)	2006	Retrospective	OAG	55	18.4	3.4	36 mo
Shoji et al. (2007)	2007	Retrospective	OAG	35	16.7	1.4	36 mo
Mierzejewski et al. (2008)	2008	Retrospective	OAG	52	18.4	4.8	14.4 +/- 4.1 mo
Fea et. al (2010)	2010	Prospective	OAG	21	17.3	3.0	15 mo
Samuelson et al. (2011)	2011	Prospective	OAG	117	18.0	3.0	12 mo
Arthur et al. (2014)	2014	Retrospective	OAG	37	16.2	4.6	24 mo
Pfeiffer et al. (2015)*	2015	Prospective	OAG	50	26.6	4.2	24 mo
Siegel et al. (2015)	2015	Retrospective	OAG	52	17.7	4.4	36 mo
Vold et al. (2016)*	2016	Prospective	OAG	131	24.5	3.0	24 mo
Jacobi et al. (1999)	1999	Prospective	PXG	22	32.0	7.7	24 mo
Damji et al. (2006)	2006	Prospective	PXG	29	19.8	2.9	24 mo
Mierzejewski et al. (2008)	2008	Retrospective	PXG	23	21.0	3.5	15.1 +/- 2.1 mo
Shingleton et al. (2008)	2008	Retrospective	PXG	51	18.0	4.0	60 mo

```
df_ <- df %>% filter(!is.na(LastPeriodAbsIOPChangeMean)) %>%
  arrange(subtype, Year, study.name) %>%
  dplyr::select(
    study.name,
    PostEyes=LastPeriodEyes,
    PostIOPMean=LastPeriodIOPMean,
    PostIOPSD=LastPeriodIOPStdDev,
    PostIOPChangeMean=LastPeriodAbsIOPChangeMean,
    PostIOPChangeSD=LastPeriodAbsIOPChangeStdDev)
kable(df_, digits = 1)
```

study.name	PostEyes	PostIOPMean	PostIOPSD	PostIOPChangeMean	PostIOPChangeSD
Hayashi et al. (2001)†	72	14.5	2.6	-7.2	3.5
Mierzejewski et al. (2008)	25	14.4	4.4	-5.1	4.8
Tham et al. (2008)	35	14.5	3.1	-1.8	3.7
Tham et al. (2009)	27	16.1	4.1	-8.3	6.3

study.name	PostEyes	PostIOPMean	PostIOPSD	PostIOPChangeMean	PostIOPChangeSD
Liu et al. (2011)	30	12.6	2.3	-3.8	4.0
Tham et al. (2013)	20	15.9	3.9	-8.4	6.0
Dias-Santos (2015)	15	14.5	1.5	-5.4	8.0
Azuara-Blanco et al. (2016)**	182	16.6	3.5	-12.9	8.0
Lee et al. (2016)	56	12.7	2.5	-1.7	4.6
Lam et al. (2008)	30	12.6	1.9	-47.1	8.4
Lee et al. (2010)	26	13.2	2.8	-35.8	10.0
Husain et al. (2012)	18	12.9	4.0	-44.5	16.2
Kim et al. (1999)	31	15.2	2.9	-2.9	3.6
Hayashi et al. (2001)†	50	15.2	3.8	-5.3	4.8
Leelachaikul et al. (2005)	54	15.2	3.7	-1.6	4.2
Mathalone et al. (2005)	24	15.1	3.2	-1.9	4.9
Damji et al. (2006)	24	NA	NA	-1.5	4.2
Shingleton et al. (2006)	55	NA	NA	-1.4	3.3
Shoji et al. (2007)	20	15.6	3.4	-1.0	3.0
Mierzejewski et al. (2008)	52	14.4	4.0	-4.0	4.0
Fea et. al (2010)	21	15.7	1.1	-1.6	3.2
Samuelson et al. (2011)	123	NA	NA	-1.0	3.3
Arthur et al. (2014)	17	14.1	4.0	-2.1	5.2
Pfeiffer et al. (2015)*	47	19.2	4.7	-7.4	5.4
Siegel et al. (2015)	52	15.5	3.6	-2.2	4.8
Vold et al. (2016)*	116	19.3	3.3	-5.4	3.9
Jacobi et al. (1999)	13	18.0	1.3	-14.0	7.4
Damji et al. (2006)	27	NA	NA	-3.1	4.1
Mierzejewski et al. (2008)	23	14.4	3.0	-6.6	3.9
Shingleton et al. (2008)	51	16.9	3.4	-1.1	4.5

12 & 15. Risk of bias

A major source of bias is loss of follow-up. Retrospective and prospective studies have different potential patterns of lossiness. Most retrospective studies reported the IOP before and after surgery *for the same set of eyes*, which leaves the possibility that some eyes received surgery and were lost in follow up, which is very hard to quantify. Prospective studies are less problematic in this regard, as the number of eyes lost in follow-up is known.

In the main text, we present the results of the analysis under a missing completely at random (MCAR) assumption, which does do not attempt to correct for these biases.

In a supplementary analysis, we verify the robustness of the results by considering the results of including just the prospective studies, which are less subject to bias. We also perform a sensitivity analysis, where we estimate how our results would change under an MNAR (missing not at random) scenario where eyes with the worst outcomes are lost to follow up.

Specifically, we consider what would happen if the eyes lost in follow-up had 0 change in IOP compared to the pre-period. We also considered what would happen if the change in IOP in the group lost in follow-up was 5 mm Hg higher than in the observed group (ref Metaanalysis an introduction with R).

Because the loss of follow-up in the prospective group was rather mild (table below), the results were not very sensitive to this form of loss, except in the case of PXG, where only two prospective studies have been performed, and they both had significant loss in follow up.

```
df <- read.data()
df <- filter.data(df, 'prospective')

df_ <- df %>% filter(subtype != 'acute', !is.na(LastPeriodEyes))
frac <- (df_$PreOpEyes - df_$LastPeriodEyes)/ df_$PreOpEyes
frac.range <- data.frame("fraction with loss"=mean(frac > 0),
           "mean loss"=mean(frac),
           "SD loss"=sd(frac),
           "min loss"=min(frac),
           "max loss"=max(frac),
```

```
            check.names=FALSE)
kable(frac.range, digits=2)
```

fraction with loss	mean loss	SD loss	min loss	max loss
0.45	0.07	0.12	-0.05	0.41

Small studies often show larger magnitude effects because of the so-called file drawer problem: the tendency of negative results from small-scale studies to never get published (ref meta-analysis book). We used standard funnel plot based methods to check if this was the case. In fact, the data showed an unusual, opposite trend: the larger studies tended to show the largest effects, an effect we attribute to washout. We discuss this in detail in the main results.

A final, important source of bias is the lack of a true control group. It would very difficult to create a true control – sham surgery – and unethical to not subject patients to any treatments. We addressed this concern by aggregating information from Stage II trials and from prevention studies to get a very coarse estimate of how large this effect might be (refs) - the estimates range from 1 to 2.5 mm Hg. This estimate is necessarily very crude, because the
test populations are not comparable (ref improper metaanalysis in glaucoma), but such an estimate, no matter how coarse, is important. In particular, the estimate should be taken into account when interpreting the extant data, and also for preparing a power analysis in future studies.

14. Synthesis of results

We used standard random-effects meta-analysis for continuous outcomes (ref `meta` package in R; ref book) throughout the text.

15. Additional analyses

In our main metaanalysis, we estimated the change in IOP that followed phacoemulsification. An ideal analysis would instead estimate the *causal* effect of surgery. The causal effect of surgery is the change in IOP following the intervention, holding all other variables fixed, compared to what would have happened if we did not perform surgery (ref causal impact). In a complementary analysis, we estimated this causal effect.

Following surgery, glaucoma medications are frequently adjusted; an ideal experiment would hold medications fixed. Studies with washout in the pre- and post- periods - temporary interruption of medication - offer an excellent approximation to the ideal fixed medication experiment.

For the studies without washout, we estimated a net effect of surgery: the sum of the measured effect in IOP and the expected change on IOP if medications has been fixed to their number in the pre-surgery period. Because we did not know the exact mix of medications used in each study, we used timolol as our reference. Timolol is the most commonly prescribed glaucoma medication, and in a recent meta-analysis, it showed an efficacy close to the mean of all tested medications (3.8 mmHg; ref).

In practice, if a study without washout reported a change in IOP of -2 mmHg, and a mean decrease in numbers of meds of 1 in the post-period, we attributed a net decrease of 5.8 mm Hg in IOP. For a study with washout in the pre- and post-periods, the attributed net change in IOP was simply the reported change in IOP.

To this net effect, we must subtract the change in IOP that would have occured in a hypothetical control arm. As outlined in the *Risk of bias* section, we do not have great estimates of this quantity. In the graph for this analysis, we indicate our best guess for this quantity (-2 mmHg), based on the extant data.

We believe this is a conservative estimate of the net causal impact of surgery for three reasons:

- Baseline: in studies without washout, the starting baseline was generally lower because of the use of drugs to control IOP. It's likely that surgery would show a larger effect on IOP when starting from a larger baseline.
- Substitution of drugs: studies did not report the mix of medications they used on patients. It's possible that more efficacious drugs that were less well tolerated were substituted for less powerful drugs with better side effect profiles once IOP was relatively well controlled.
- Intent-to-treat: in the real world, adherence for preventive drugs can be low. Surgery does not suffer such an adherence problem, so its net relative effect can be amplified compared to meds in an intent-to-treat analysis.

Other things to discuss

- Very little change across time
- Sensitivity to loss of follow up
- Ideal experiment design based on review of literature
- For PXG, there's high loss of follow up, and few good studies – more research is necessary.

Additional references

- Borenstein et al. (2009), Introduction to Meta-Analysis. [Link].
- Schwarzer et al. (2015), Meta-analysis with R. [Link].
- Kadic et al. (2016), Extracting data from figures with software was faster, with higher interrater reliability than manual extraction. [Link].
- Li & Dickersin (2013), Citation of Previous Meta-analyses on the Same Topic: A Clue to Perpetuation of Incorrect Methods? [Link].
- Li et al. (2016), Comparative Effectiveness of First-Line Medications for Primary Open-Angle Glaucoma: A Systematic Review and Network Meta-analysis. [Link].
- Imbens & Rubin (2015), Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. [Link].