Kotani et al re-analysis

Ben Prytherch

2023-04-19

Read data, get estimates and variances

propofol_metaanalysis_data.csv is original data.

propofol_metaanalysis_data2.csv changes the denominators for Likhvantsev 2016 are changed from 450 each to 396 and 292.

propofol_metaanalysis_data3.csv uses 30-day mortality values rather than 1-year values for Likhvantsev 2016: 20/431 and 17/437

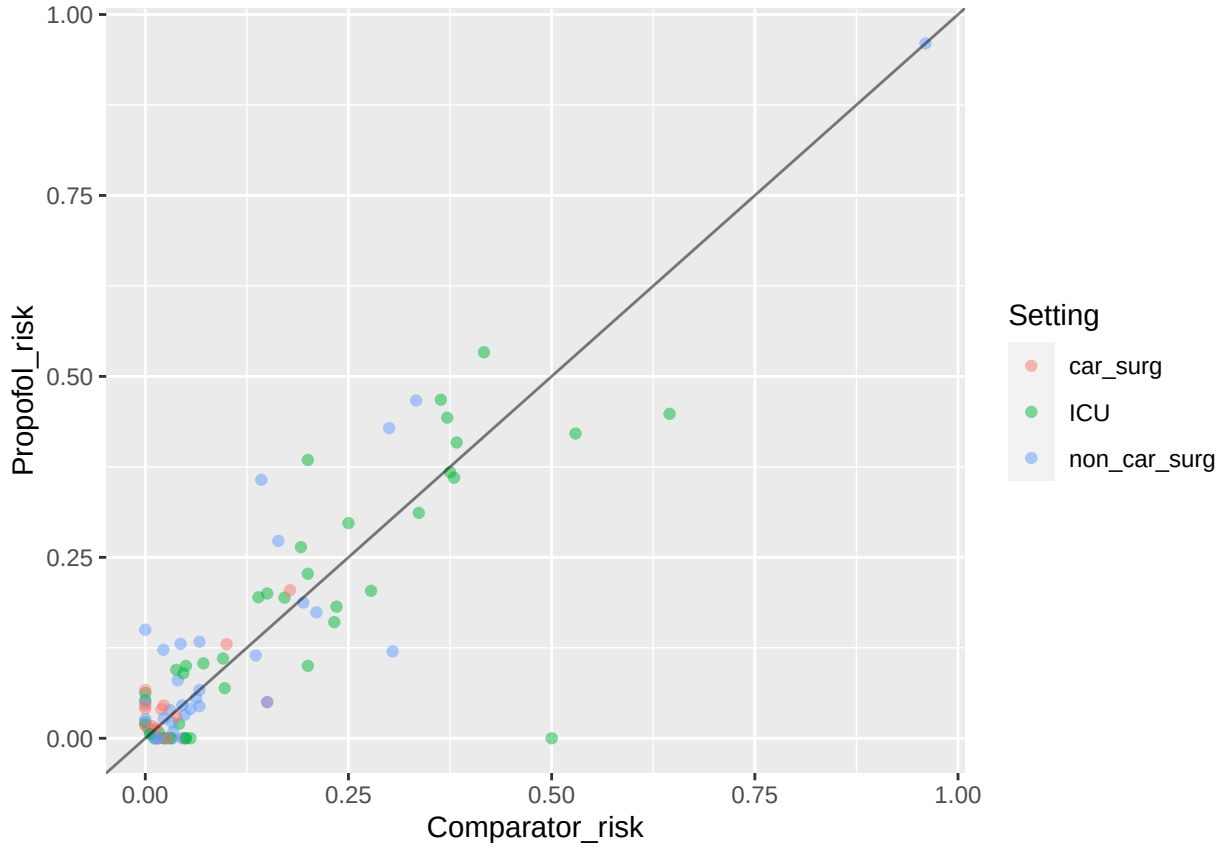
```
##propofol_data <- read.csv("propofol_metaanalysis_data.csv")
propofol_data <- read.csv("propofol_metaanalysis_data2.csv")
#propofol_data <- read.csv("propofol_metaanalysis_data3.csv")


#propofol_data <- propofol_data[-236,]
#uncomment above to remove Wu ZF 2019, 23/24 deaths in both groups

propofol_data$Propofol_risk <- propofol_data$Propofol_Events/propofol_data$Propofol_Total
propofol_data$Comparator_risk <- propofol_data$Comparator_Events/propofol_data$Comparator_Total

##All the data
propofol_es <- escalc(measure="RR",data=propofol_data,
                      ai=Propofol_Events,
                      n1i=Propofol_Total,
                      ci=Comparator_Events,
                      n2i=Comparator_Total)

##Do not estimate RR for studies with a zero count;
##(method used in Kotani et al)
propofol_es_NAs <- escalc(measure="RR",data=propofol_data,drop00 = T,
                      ai=Propofol_Events,
                      n1i=Propofol_Total,
                      ci=Comparator_Events,
                      n2i=Comparator_Total)

##Clear out the studies with zero count
##This is used for the plots below
propofol_rm_NA <- subset(propofol_es_NAs,propofol_es_NAs$yi!="NA")
```

Plot data

The scatterplot below compares propofol vs. comparator risk across studies. Outlier is Wu ZF (2009); 24/25 deaths in both groups.

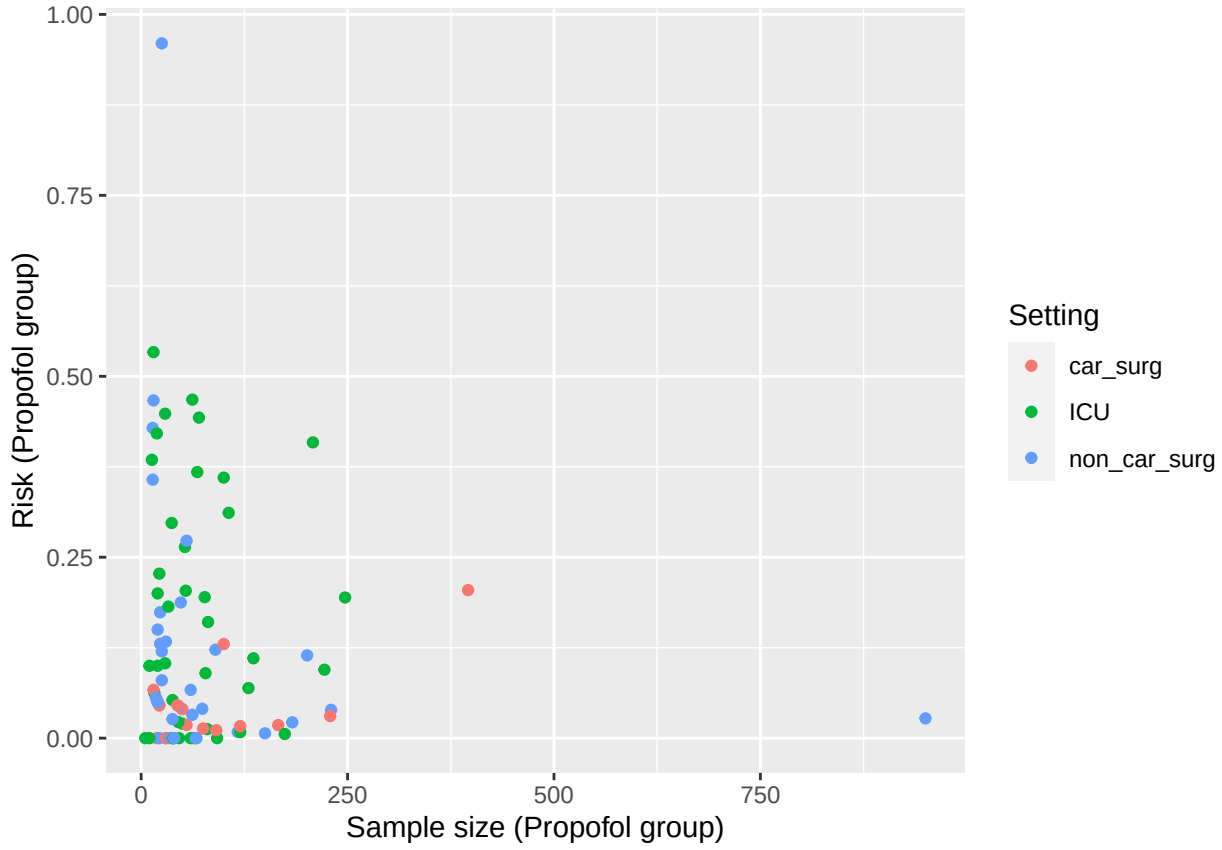
Note that, while cardiac surgery studies show larger RR, largest indivdual risks all come from ICU and non-cardiac surgery studies. Variability in RR across studies is extreme, showing large heterogeneity in overall mortality risk across studies.

```
ggplot(propofol_rm_NA,aes(x=Comparator_risk,y=Propofol_risk,col=Setting)) +
  geom_point(alpha=0.5) +
  geom_abline(slope=1,intercept=0,alpha=0.5)
```



Plot of sample size in propofol group vs propofol risk. Another way to show large heterogeneity in risk within each of the three settings, despite $I^2 = 0$ for RR analysis.

```
ggplot(propofol_rm_NA,aes(x=Propofol_Total,y=Propofol_risk,col=Setting)) +
  geom_point()+
  labs(y="Risk (Propofol group)",x="Sample size (Propofol group)")
```



Fitted models

Model zero: propofol risk only

This first model is **not** for RR; it is propofol risk only. I'm including it in case the heterogeneity statistic is desired; it's $I^2 = 0.98$, or $I^2 = 0.938$ if Wu ZF 2019 is removed.

Seems like a heroic assumption to say that "population" RR (if one accepts such an abstraction) is constant across raw risks that range from nearly 0% to above 50%.

```
propofol_risk_only <- escalc(measure="PR",data=propofol_data,
                             xi=Propofol_Events,
                             ni=Propofol_Total)

model0 <- rma(yi,vi,data=propofol_risk_only)
model0
```

```
##
## Random-Effects Model (k = 252; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.0065 (SE = 0.0007)
## tau (square root of estimated tau^2 value):      0.0804
## I^2 (total heterogeneity / total variability):   98.24%
## H^2 (total variability / sampling variability):  56.95
##
```

```
## Test for Heterogeneity:
## Q(df = 251) = 1607.6169, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.0468  0.0055  8.5059  <.0001  0.0360  0.0576  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 1: traditional random effects

This is the random-effects model fitting using REML (metafor's default).

```
model1 <- rma(yi,vi,data=propofol_rm_NA)
model1
```

```
##
## Random-Effects Model (k = 92; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0065)
## tau (square root of estimated tau^2 value):      0
## I^2 (total heterogeneity / total variability):   0.00%
## H^2 (total variability / sampling variability):  1.00
##
## Test for Heterogeneity:
## Q(df = 91) = 63.5027, p-val = 0.9874
##
## Model Results:
##
## estimate      se    zval    pval    ci.lb   ci.ub
##   0.0356  0.0363  0.9821  0.3261  -0.0355  0.1067
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 2: the one from the meta-analysis

This is the common-effect Mantel-Haenszel model used in Kotani et al.

```
model2 <- rma.mh(data=propofol_rm_NA,
                 ai=Propofol_Events,
                 n1i=Propofol_Total,
                 ci=Comparator_Events,
                 n2i=Comparator_Total)
model2
```

```
##
## Equal-Effects Model (k = 92)
##
## I^2 (total heterogeneity / total variability):  0.00%
```

```
## H^2 (total variability / sampling variability): 0.71
##
## Test for Heterogeneity:
## Q(df = 91) = 64.3905, p-val = 0.9844
##
## Model Results (log scale):
##
## estimate      se    zval    pval    ci.lb   ci.ub
##   0.0913  0.0616  1.4819  0.1384  -0.0294  0.2120
##
## Model Results (OR scale):
##
## estimate   ci.lb   ci.ub
##   1.0956  0.9710  1.2361
##
## Cochran-Mantel-Haenszel Test:    CMH =  2.1101, df = 1,  p-val = 0.1463
## Tarone's Test for Heterogeneity: X^2 = 90.7984, df = 91, p-val = 0.4862
```

Model 3: Common effect model w/ inverse variance weighting

This is the common-effect model fit using inverse-variance weighting. This just shows results are the same as in the random-effects model, as heterogeneity estimate is 0.

```
model3 <- rma(yi,vi,data=propofol_rm_NA,method="EE")
model3
```

```
##
## Equal-Effects Model (k = 92)
##
## I^2 (total heterogeneity / total variability):   0.00%
## H^2 (total variability / sampling variability):  0.70
##
## Test for Heterogeneity:
## Q(df = 91) = 63.5027, p-val = 0.9874
##
## Model Results:
##
## estimate      se    zval    pval    ci.lb   ci.ub
##   0.0356  0.0363  0.9821  0.3261  -0.0355  0.1067
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 4: add setting as a moderator

Now we fit a mixed-effects model with "setting" as a moderator.

```
propofol_rm_NA$Setting <- as.factor(propofol_rm_NA$Setting)
model4 <- rma(measure="RR",yi,vi,data=propofol_rm_NA,method="REML",
              mods = ~ Setting)
model4
```

```
##
## Mixed-Effects Model (k = 92; tau^2 estimator: REML)
##
## tau^2 (estimated amount of residual heterogeneity):     0 (SE = 0.0126)
## tau (square root of estimated tau^2 value):             0
## I^2 (residual heterogeneity / unaccounted variability): 0.00%
## H^2 (unaccounted variability / sampling variability):   1.00
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 89) = 62.4220, p-val = 0.9855
##
## Test of Moderators (coefficients 2:3):
## QM(df = 2) = 1.0807, p-val = 0.5825
##
## Model Results:
##
##                      estimate      se     zval    pval    ci.lb   ci.ub
## intrcpt                0.1647  0.1326   1.2418  0.2143  -0.0952  0.4246
## SettingICU            -0.1298  0.1437  -0.9032  0.3664  -0.4115  0.1519
## Settingnon_car_surg   -0.1479  0.1423  -1.0394  0.2986  -0.4267  0.1310
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare estimated relative risk across setting

This extracts RR estimates and CIs across setting (it could be made cleaner). Note that *predict()* returns confidence interval bounds first, then prediction interval bounds, which probably aren't of interest.

```
car_surg_est <- predict(model4,newmods=c(0,0),transf=exp,digits=2)
ICU_est <- predict(model4,newmods=c(1,0),transf=exp,digits=2)
non_car_surg_est <- predict(model4,newmods=c(0,1),transf=exp,digits=2)

estimates_by_setting <-
  rbind(print(car_surg_est),print(ICU_est),print(non_car_surg_est))
```

```
estimates_by_setting
```

```
##       pred     ci.lb    ci.ub     pi.lb    pi.ub
##   1.179036 0.9091487 1.529042 0.9091487 1.529042
## 1 1.035501 0.9289545 1.154267 0.9289545 1.154267
## 2 1.016956 0.9193222 1.124959 0.9193222 1.124959
```

Same again, but on log scale. Perhaps useful if anyone focuses in on how far the upper bounds of the CIs on RR scale extend compared to the lower bounds.

```
car_surg_est_log <- predict(model4,newmods=c(0,0),digits=2)
ICU_est_log <- predict(model4,newmods=c(1,0),digits=2)
non_car_surg_est_log <- predict(model4,newmods=c(0,1),digits=2)

log_estimates_by_setting <-
  rbind(print(car_surg_est_log),print(ICU_est_log),print(non_car_surg_est_log))
```

```
log_estimates_by_setting
```

```
##         pred         se       ci.lb     ci.ub       pi.lb     pi.ub
##   0.16469750 0.13262700 -0.09524663 0.4246416 -0.09524663 0.4246416
## 1 0.03488501 0.05539927 -0.07369557 0.1434656 -0.07369557 0.1434656
## 2 0.01681391 0.05149715 -0.08411865 0.1177465 -0.08411865 0.1177465
```

##Including the studies with zero counts

This uses metafor's default of adding 0.5 to mortality counts in both propofol and comparator if either is zero. The results is that RR among cardiac surgery studies goes down slighly from 1.18 (0.91, 1.53) to 1.16 (0.91, 1.48), is effectively unchanged among ICU studies, and goes up very slightly among non-cardiac surgery studies.

```
propofol_es$Setting <- as.factor(propofol_es$Setting)
model5 <- rma(measure="RR",yi,vi,data=propofol_es,method="REML",
              mods = ~ Setting)
model5
```

```
##
## Mixed-Effects Model (k = 252; tau^2 estimator: REML)
##
## tau^2 (estimated amount of residual heterogeneity):     0 (SE = 0.0115)
## tau (square root of estimated tau^2 value):             0
## I^2 (residual heterogeneity / unaccounted variability): 0.00%
## H^2 (unaccounted variability / sampling variability):   1.00
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 249) = 67.4676, p-val = 1.0000
##
## Test of Moderators (coefficients 2:3):
## QM(df = 2) = 0.8642, p-val = 0.6492
##
## Model Results:
##
##                      estimate      se     zval    pval    ci.lb   ci.ub
## intrcpt                0.1480  0.1242   1.1915  0.2335  -0.0955  0.3915
## SettingICU            -0.1133  0.1359  -0.8331  0.4048  -0.3797  0.1532
## Settingnon_car_surg   -0.1238  0.1338  -0.9252  0.3549  -0.3859  0.1384
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
car_surg_est_zeroes <- predict(model5,newmods=c(0,0),transf=exp,digits=2)
ICU_est_zeroes <- predict(model5,newmods=c(1,0),transf=exp,digits=2)
non_car_surg_est_zeroes <- predict(model5,newmods=c(0,1),transf=exp,digits=2)

estimates_by_setting_zeroes <- rbind(print(car_surg_est_zeroes),print(ICU_est_zeroes),print(non_car_surg
estimates_by_setting_zeroes
```

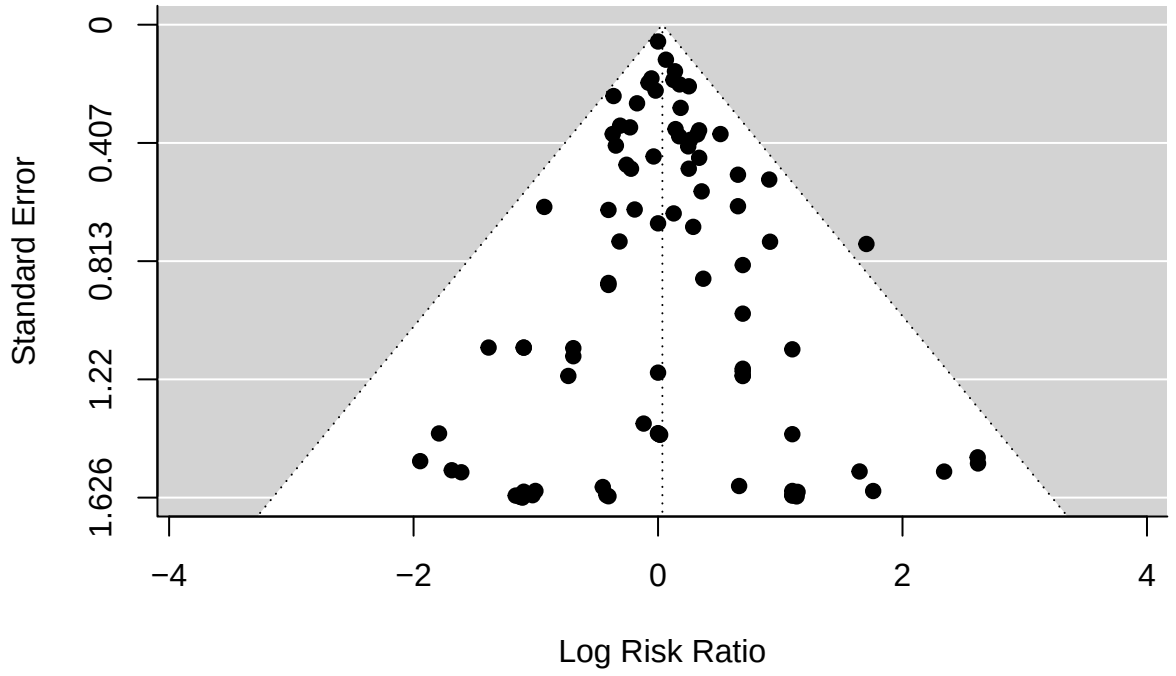
```
estimates_by_setting_zeroes
```

```
##       pred     ci.lb    ci.ub     pi.lb    pi.ub
##   1.159542 0.9089415 1.479236 0.9089415 1.479236
## 1 1.035386 0.9292837 1.153604 0.9292837 1.153604
## 2 1.024567 0.9297196 1.129091 0.9297196 1.129091
```

Publication bias checks

This reproduces the funnel plot we've already seen. Visually there are a handful slightly larger right sided value than left sided values, but if anything is there it's mild. Formal tests find nothing. This doesn't seem like a concern.

```
funnel(model1)
```




```
regtest(model1)
```

```
##
## Regression Test for Funnel Plot Asymmetry
##
## Model:     mixed-effects meta-regression model
## Predictor: standard error
##
## Test for Funnel Plot Asymmetry: z = 0.3414, p = 0.7328
## Limit Estimate (as sei -> 0):   b = 0.0251 (CI: -0.0680, 0.1183)
```

```
##Not worth printing
#trimfill(model1,side= "right")
#trimfill(model1,side = "left")
```

My takeaways and comments

- Fixing Likhvantsev 2016 has a large impact on fixed-effects estimate and CI. It also has a large impact on random-effects estimate and CI *within cardiac surgery setting*. Unsurprising, given its sample size was very big relative to the other cardiac surgery studies.
- Heterogeneity in overall mortality risk is extreme.
- Aggregating across dissimilar comparisons to propofol to create a single "comparator" group sounds bonkers to me, but I have zero domain knowledge here. Adrian Simpson's 2018 "Princesses are bigger than elephants: effect size as a category error in evidence based education" does an excellent job of pointing out the absurdity of such aggregation (he's in education research).
- *"All models are wrong, but some are wronger"*. Common effect model makes a wildly implausible assumption, but random effects model isn't "correct" either. "True" RR is a mathematical abstraction. To the extent it represents reality well, its natural log is not normally distributed across study settings. RRs differ across studies for all kinds of reasons; meta-analysis inevitably requires averaging over variables that shouldn't be averaged over. We do it because the alternative is unyielding complexity.

  One of the co-authors has said something along the lines of "even if these results aren't statistically significant, they lean in the direction of increased risk from propofol and so they deserve attention."

  I dislike statistical signficance testing and am normally sympathetic to this kind of argument; $p < 0.05$ is arbitrary, the Type II error rate is never known, and the point null is usually implausible. However, "non-significant but in the predicted direction" is not so compelling when there are so many unmeasured moderators at play. Variance in study outcomes is not i.i.d. noise, and for me this makes "almost significant" particularly unpersuasive.