Focused model selection for meta-analysis of 2×2 tables

Céline Cunen

University of Oslo

10/05/2016

The Problem - Meta-analysis of 2x2 tables

Study	Sample size control $= m_{0,i}$	Events control $= y_{0,i}$	Sample size treatment $= m_{1,i}$	Events treatment $= y_{1,i}$
1	39	1	43	2
2	44	4	44	4
3	107	4	110	6
4	103	5	100	7
5	110	3	106	7
6	154	4	146	11

- We are interested in finding out whether the control and treatment groups are different
- Thus, we have a natural focus parameter, the odds ratio
- Many methods/estimators in the literature, but choice of model is seldom made explicit

Binomial Model

$$Y_{0,i} \sim \mathrm{Bin}(m_{0,i}, p_{0,i}) \quad \text{and} \quad Y_{1,i} \sim \mathrm{Bin}(m_{1,i}, p_{1,i})$$

- with $p_{0,i} = \frac{e^{\theta_i}}{1+e^{\theta_i}}$ and $p_{1,i} = \frac{\gamma_i e^{\theta_i}}{1+\gamma_i e^{\theta_i}}$
- Odds ratio: $\mathrm{OR}_i = \frac{p_{1,i}/(1-p_{1,i})}{p_{0,i}/(1-p_{0,i})} = \frac{\gamma_i e^{\theta_i}}{e^{\theta_i}} = \gamma_i$

Binomial Model - fixed effect

$$Y_{0,i} \sim \mathrm{Bin}(m_{0,i}, p_{0,i}) \quad \text{and} \quad Y_{1,i} \sim \mathrm{Bin}(m_{1,i}, p_{1,i})$$

- With $p_{0,i} = \frac{e^{\theta_i}}{1+e^{\theta_i}}$ and $p_{1,i} = \frac{\gamma e^{\theta_i}}{1+\gamma e^{\theta_i}}$
- Odds ratio: $\mathrm{OR} = \frac{p_{1,i}/(1-p_{1,i})}{p_{0,i}/(1-p_{0,i})} = \frac{\gamma e^{\theta_i}}{e^{\theta_i}} = \gamma$
- Usual assumption: a common odds ratio (γ) across tables

Binomial Model - natural estimators

- Maximum likelihood: $\hat{\gamma}_{ML}$, but problematic
- Conditional maximum likelihood: $\hat{\gamma}_b$

Let $z_i = y_{0,i} + y_{1,i}$, and we can consider the distribution of $y_{1,i} \mid z_i$ (excentric hypergeometric):

$$f(y_{1,i} \mid z_i) = \frac{\binom{m_{0,i}}{z_i - y_{1,i}} \binom{m_{1,i}}{y_{1,i}} \gamma^{y_{1,i}}}{\sum_{u=0}^{z_i} \binom{m_{0,i}}{z_i - u} \binom{m_{1,i}}{u} \gamma^u}, \text{ for } y_{1,i} \text{ from 0 to } \min(z_i, m_{1,i})$$

and find $\hat{\gamma}_b$ by maximising $\sum \log f(y_{1,i} \mid z_i)$.

Poisson Model

$$Y_{0,i} \sim \mathrm{Pois}(e_{0,i}\lambda_i) \quad \text{and} \quad Y_{1,i} \sim \mathrm{Pois}(e_{1,i}\lambda_i\gamma_*)$$

$e_{0,i}$ and $e_{1,i}$ are exposure weights, with $e_{0,i} = m_{0,i}/100$ and $e_{1,i} = m_{1,i}/100$

- the two models are used interchangeably when the event probabilities are small, for one table:

$$\gamma_i = \frac{p_{1,i}/(1-p_{1,i})}{p_{0,i}/(1-p_{0,i})} \approx \frac{p_{1,i}}{p_{0,i}} \approx \frac{e_{1,i}\lambda_i\gamma_{i,*}/m_{1,i}}{e_{0,i}\lambda_i/m_{0,i}} = \frac{\lambda_i\gamma_{i,*}}{\lambda_i} = \gamma_{i,*}.$$

- $\lambda_i \approx$ the risk in the control group
- γ_*: a least false parameter

Poisson Model - natural estimator

- The maximum likelihood estimator and the conditional maximum likelihood estimator are equal: $\hat{\gamma}_p$
- Here, conditioning on the total number of events in each study (z_i) gives a binomial distribution:

$$Y_{1,i} \,|\, (Z_i = z_i) \sim \mathrm{Bin}\left(z_i, \frac{e_{1,i}\gamma_*}{e_{0,i} + e_{1,i}\gamma_*}\right)$$

- When $m_{0,i} = m_{1,i}$, $\hat{\gamma}_p = \frac{\sum y_{1,i}}{\sum y_{0,i}}$
- A machine producing estimates, not the true model

Model selection scheme

We assume that the binomial model is the true one, and "evaluate" the two estimators with respect to that model. Classical FIC (with local misspecification) is not applicable.

FIC

$$\mathrm{FIC}_b = \widehat{\mathrm{mse}}(\hat{\gamma}_b) = \widehat{\mathrm{Var}(\hat{\gamma}_b})$$
$$\mathrm{FIC}_p = \widehat{\mathrm{mse}}(\hat{\gamma}_p) = \widehat{b^2} + \widehat{\mathrm{Var}(\hat{\gamma}_p})$$

Risk formulae

For this simple case (fixed effect models) we can work out formulae for the limiting risk. For the estimator from the binomial model:

$$\mathrm{Var}(\hat{\gamma}_b) \doteq \frac{\gamma^2}{\sum \mathrm{E}(\mathrm{Var}(Y_{1,i} \mid z_i))},$$

where $\mathrm{Var}(Y_{1,i} \mid z_i)$ is the variance of the excentric hypergeometric, and the expectation is taken over the distribution of z_is.
For the estimator from the poisson model we have:

$$b \doteq \mathrm{E}(\hat{\gamma}_b) - \mathrm{E}(\hat{\gamma}_p) = \gamma - \gamma_*$$

with γ_* the solution to

$$\frac{\sum m_{1,i} p_{1,i}}{\gamma_*} = \sum \frac{m_{1,i}(m_{0,i} p_{0,i} + m_{1,i} p_{1,i})}{m_{0,i} + m_{1,i}\gamma_*},$$

and

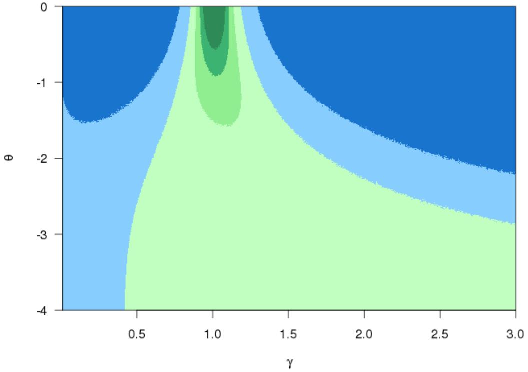
$$\mathrm{Var}(\hat{\gamma}_p) \doteq (1/k)\frac{K}{J^2} = \frac{\sum_{i=1}^k \left\{ \frac{m_{0,i}^2 m_{1,i} p_{1,i}(1-p_{1,i})}{\gamma_*^2 (m_{0,i}+m_{1,i}\gamma_*)^2} + \frac{m_{1,i}^2 m_{0,i} p_{0,i}(1-p_{0,i})}{(m_{0,i}+m_{1,i}\gamma_*)^2} \right\}}{\left[\sum_{i=1}^k \left\{ -\frac{m_{1,i} p_{1,i}}{\gamma_*^2} + \frac{m_{1,i}^2 (m_{0,i} p_{0,i} + m_{1,i} p_{1,i})}{(m_{0,i}+m_{1,i}\gamma_*)^2} \right\} \right]^2}$$

Risk ratio: where does the Poisson model win?

For $m_0 = m_1 = 30$, $k = 30$ and θs drawn from a normal with variance 0.1^2, we look at $\mathrm{mse}_b/\mathrm{mse}_p$.

Green area: estimator from Poisson model wins.

Blue area: estimator from binomial model wins.

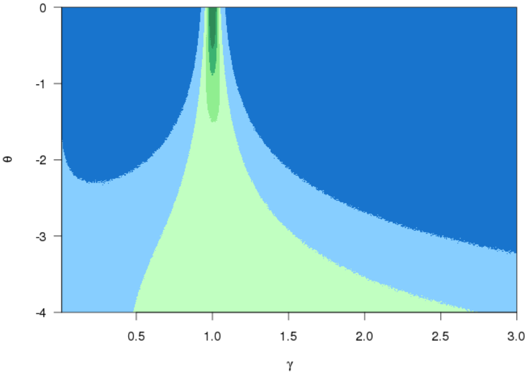


Risk ratio: where does the Poisson model win?

For $m_0 = m_1 = 300$, $k = 30$ and θs drawn from a normal with variance 0.1^2, we look at $\mathrm{mse}_b / \mathrm{mse}_p$.

Green area: estimator from Poisson model wins.

Blue area: estimator from binomial model wins.



FIC - estimated risk

The variances:

$$\widehat{\mathrm{Var}}(\hat{\gamma}_b) = \hat{J}_b^{-1}$$
$$\widehat{\mathrm{Var}}(\hat{\gamma}_p) = (1/k)\frac{\hat{K}_p}{\hat{J}_p^2},$$

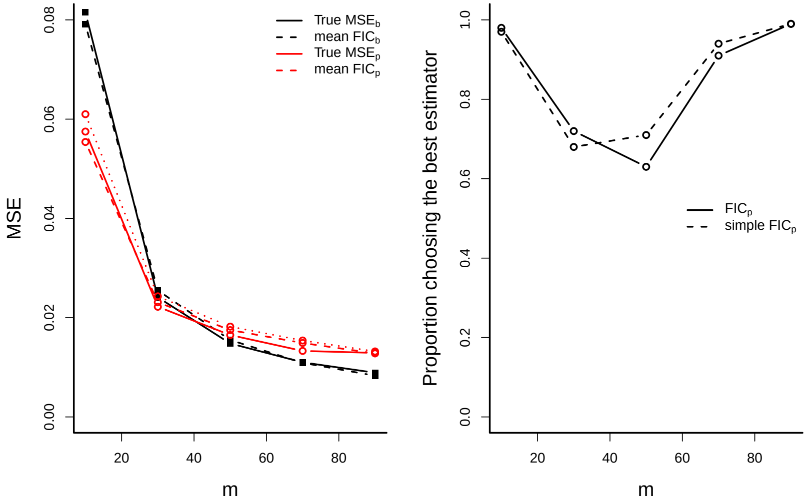
with $\hat{J}_b^{-1}$ from the conditional binomial model, and $\hat{J}_p^2$ and $\hat{K}_p$ from the conditional Poisson model.

The bias:

- $\hat{b}^2 = (\hat{\gamma}_p - \hat{\gamma}_b)^2$ over-estimates the squared bias
- Better estimate: $\hat{b}^2 = (\hat{\gamma}_p - \hat{\gamma}_b)^2 - \widehat{\mathrm{Var}}(\hat{b})$
- Where we estimate $\mathrm{Var}(\hat{b})$ by parametric bootstrap

FIC - does it work?

Simulations with $k = 50$, $\gamma = 1.5$ and θs drawn from a normal with mean -2 and variance 0.1^2.



Real example - small event probabilities

For the Lidocaine data, with $k = 6$, $\bar{m} = 92$ and (naively) estimated event probabilities of $\hat{p}_0 = 0.05$ and $\hat{p}_1 = 0.09$, I get

	$\hat{\gamma}$	$\hat{b}^2$	$\widehat{\mathrm{Var}}(\hat{\gamma})$	$\sqrt{FIC}$
Conditional binomial	1.783	0	0.249	0.499
Poisson	1.733	0.00..	0.063	0.251

Estimator from Poisson model is best!

Real example - larger event probabilities

For the Catheter data, with $k = 13$, $\bar{m} = 101$ and (naively) estimated event probabilities of $\hat{p}_0 = 0.28$ and $\hat{p}_1 = 0.16$, I get

	$\hat{\gamma}$	$\hat{b}^2$	$\widehat{\mathrm{Var}}(\hat{\gamma})$	$\sqrt{FIC}$
Conditional binomial	0.440	0	0.0020	0.0447
Poisson	0.555	0.0131	0.0047	0.1334

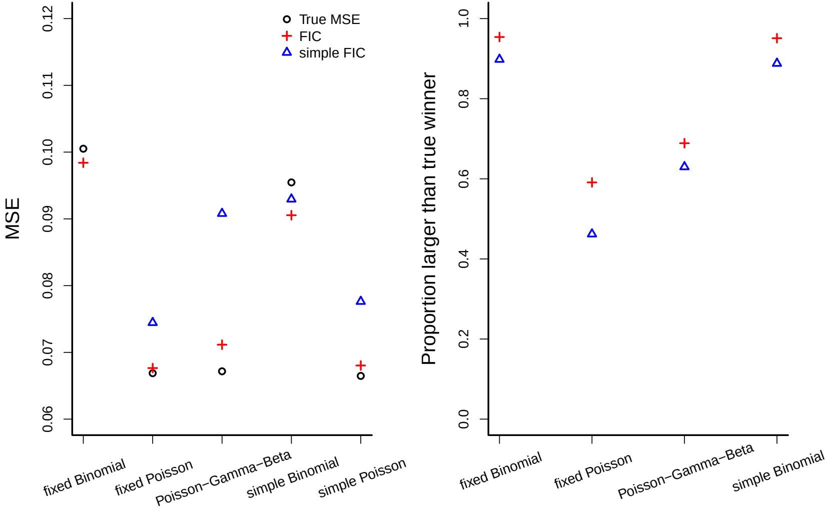
Estimator from binomial model is best!

Extensions to more models

- Still assuming the fixed effect binomial model is the "truth", we can include more models to compare
- For some models explicit risk formulas are harder to work out, but FIC can be computed by finding the "sandwich" matrix and estimating the squared bias in the same way as for the Poisson model
- Models to consider:
  - The simplest binomial model with 2 parameters
  - The simplest Poisson model with 2 parameters
  - Different kinds of random effect models, for example the Poisson-Gamma-Beta model with 4 parameters (Cai et al. 2010)

Example with more models

Simulations with $k = 30$, $\bar{m}_0 = \bar{m}_1 = 15$, $\gamma = 1.5$ and θs drawn from a normal with mean -2 and variance 0.1^2.



Conclusions

- Using an estimator from the Poisson model is appropriate for many meta-analysis situations
- The estimator from the Poisson model has a bias, but can have much smaller variance than the binomial estimator
- Further work:
  - A wider true model (with $k + 2$ parameters)
  - Other focus parameters, for example the risk difference
  - Include covariates