UMC Utrecht

Publication bias tests for survival data: a simulation study

Thomas P.A. Debray, K.G.M. Moons, H. Koffijberg, R.D. Riley

What is publication bias?

"Meta-analyses based on published group data may be affected by the selective nonpublication or late publication of negative findings" - Ioannidis et al, Journal of Clinical Epidemiology (1999)

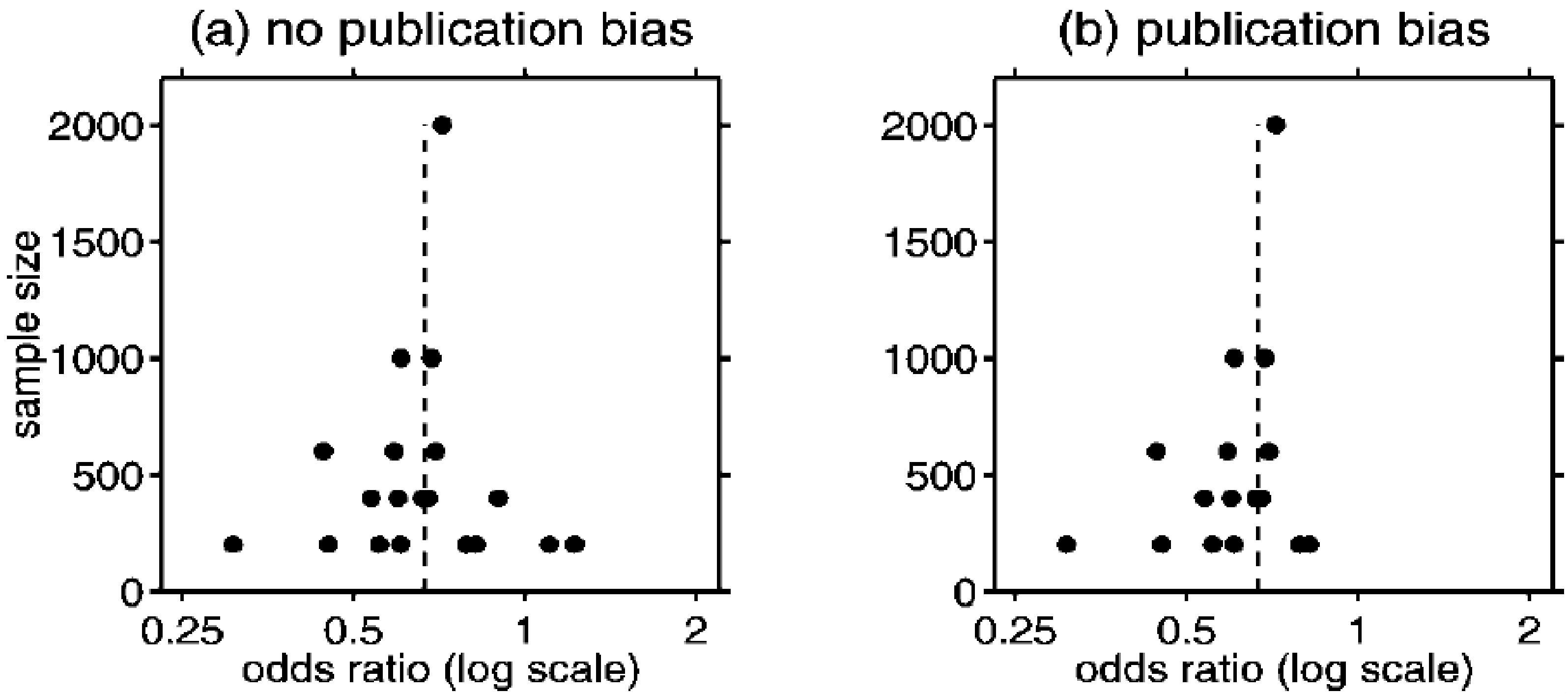
"The tendency for increased publication rates among studies that show a statistically significant effect of treatment" - Sterne et al, Journal of Clinical Epidemiology (2000)

"This bias arises when the published studies identified for inclusion in the meta-analysis do not represent all studies on the topic of interest" - Macaskill et al, Statistics in Medicine (2001)

Visual detection of publication bias

Funnel plot

- Plot reported treatment effects against their precision
- Evaluate the presence of asymmetry



Statistical testing of publication bias

- Test for asymmetry in the funnel plot
  - treatment effect: log OR or log HR
  - precision: based on standard error or sample size
- Weighted regression analysis
  - treatment effect versus standard error (**Egger**)
  - treatment effect versus sample size (**Macaskill**)
  - treatment effect versus inverse of sample size (**Peters**)

with weight = 1/(error variance of treatment effect) to allow for possible heteroscedasticity

Detecting publication bias: problems

- Asymmetry in funnel plot may arise for reasons beyond publication bias
  - log of OR (or HR) and its SE are correlated
    we should use sample size as measure of precision
- However, sample size does not well reflect precision for treatment effects resulting from survival data
  (as only events contribute to the likelihood)

To what extent can existing methods be used (i.e. Egger, Macaskill & Peters) when investigating the presence of publication bias in survival studies?

A new test for detecting publication bias

- Adopt approach Peters
- Replace sample size by total number of events (**T1**)

$$\hat{\beta}_k = a + b\,\frac{1}{\boldsymbol{d}_k} + \boldsymbol{\epsilon}_k \quad \text{weighted by } \mathbf{w}_k = \mathrm{SE}(\hat{\beta}_k)^{-2}$$

- Replace sample size by total follow-up time (**T2**)

$$\hat{\beta}_k = a + b\,\frac{1}{\boldsymbol{z}_k} + \boldsymbol{\epsilon}_k \quad \text{weighted by } \mathbf{w}_k = \mathrm{SE}(\hat{\beta}_k)^{-2}$$

Simulation study

Compare the following methods

- Egger, Macaskill, Peters
- Novel tests (T1 and T2)

Investigate influence of

- size of the meta-analysis (m)
- size of true effect (β)
- proportion of censored events (π)
- between-study heterogeneity in baseline hazard (τ)

Outcomes of interest

- Power
- Type-I error rate

Simulation study

Data generation model

- Weibull distributed survival times
- Allow for heterogeneous baseline hazard
- Allow for non-informative right-censoring

$$\mathbf{T}_{ij} = \left(- \frac{\log(u)}{l_j \exp(\beta \mathbf{X}_{ij})} \right)^{\frac{1}{v}}$$

$$l_j \sim \Gamma\left(\lambda, \tau_\lambda^2\right)$$

Statistical analysis model

- Cox regression (HR as measure of treatment effect)

Simulation study

Test statistics for publication bias (PB)

- **Type-I error**: if there is no PB mechanism, how often does the test incorrectly identifies PB presence?
- **Power**: if there is a PB mechanism, how often does the test correctly identifies PB presence?
  PB is introduced based on one-sided p-value of HR

P-value HR	P published	
	Moderate PB	**Severe PB**
< 0.05	1	1
0.05 – 0.20	0.75	0.75
0.20 – 0.50	0.50	0.25
> 0.50	0.27	0.25

Simulation study

2 types of publication bias (PB)

- **Implicit:** is there an underlying mechanism of PB?
  The current PB mechanism allows:
  - some meta-analyses to publish all generated studies.
  - some meta-analyses to publish none of the generated studies
  - one or more studies to remain unpublished for a certain number of published studies
- **Explicit:** in the current set of published studies, are some studies missing?
  - Only conduct PB test if at least 3 studies are available

Focus on detecting implicit PB here
(estimates of power will be pessimistic)

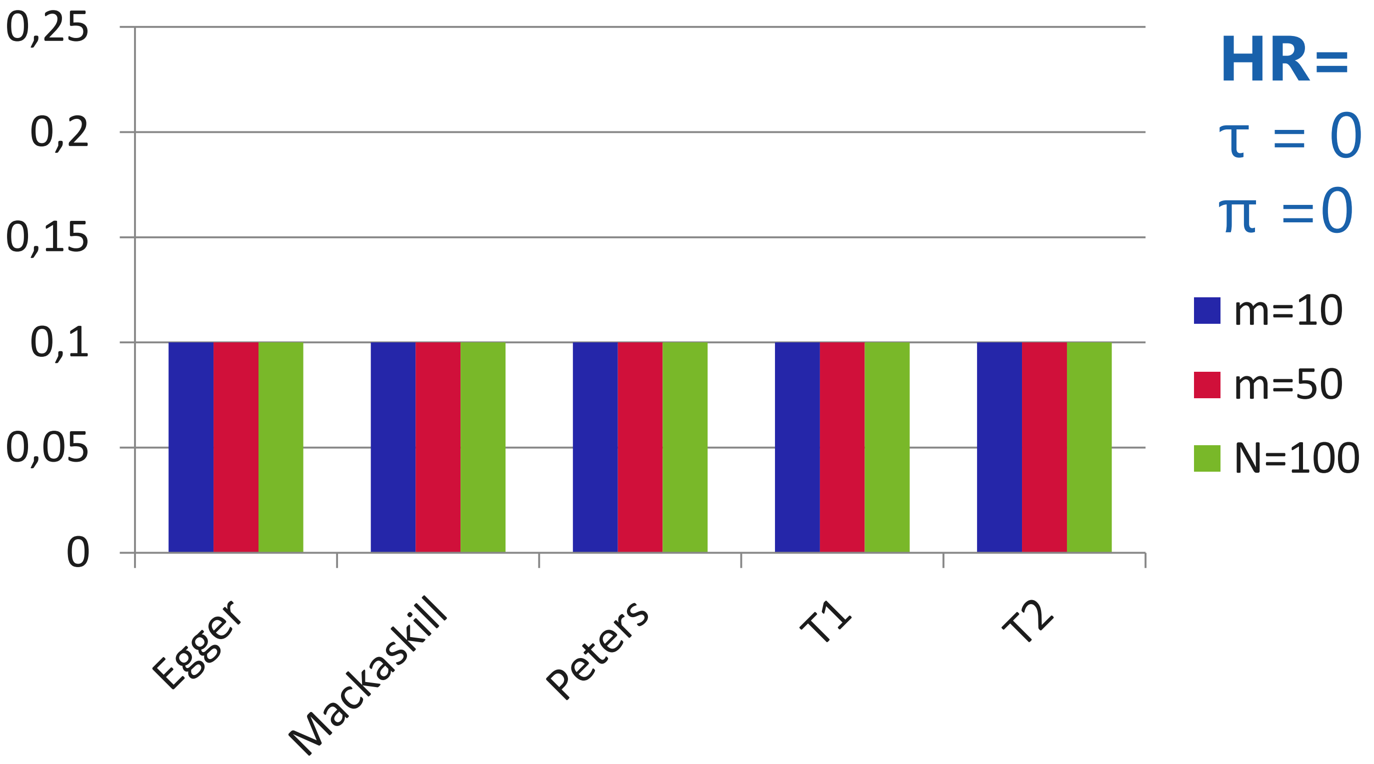
Simulation study

Setup parameters

- Generate 100,000 meta-analyses for each scenario.
- size meta-analysis: m=10, 20, 50 and 100
- sample size individual studies: generated from log-normal distribution (N=100 to 1628 for 1-99 quantile)
- Baseline hazard and rate in accordance to trial data Hodgkin's disease (parameter has value 0.03)
- true effect size $\log(\beta) = 5, 1, 0.75, 0.5$
- censoring proportion $\pi = 0, 0.30$
- heterogeneity in baseline hazard
  $\tau = 0$ (none), 0.01 (moderate), 0.02 (severe)

Simulation results: Type-I error

Seems to look good…



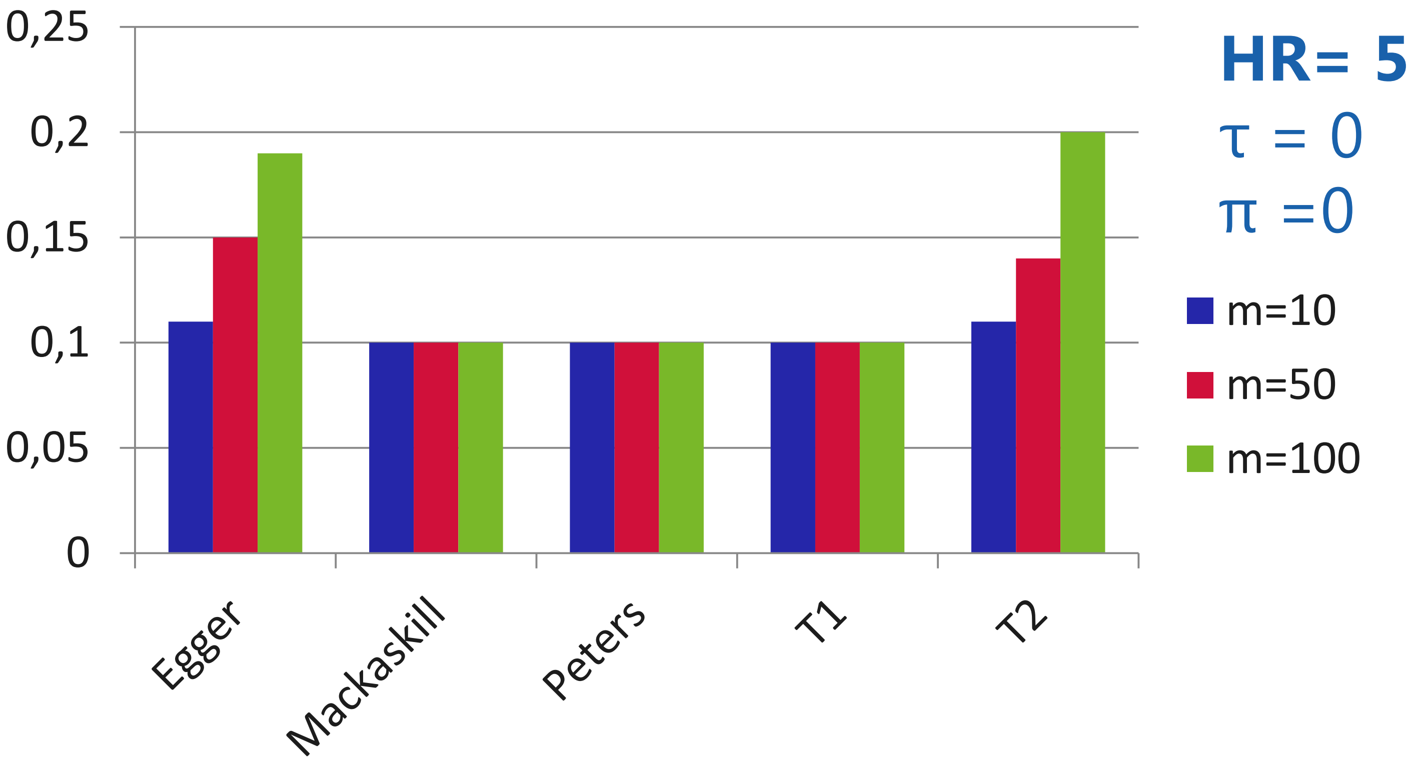
HR= 1

τ = 0

π =0

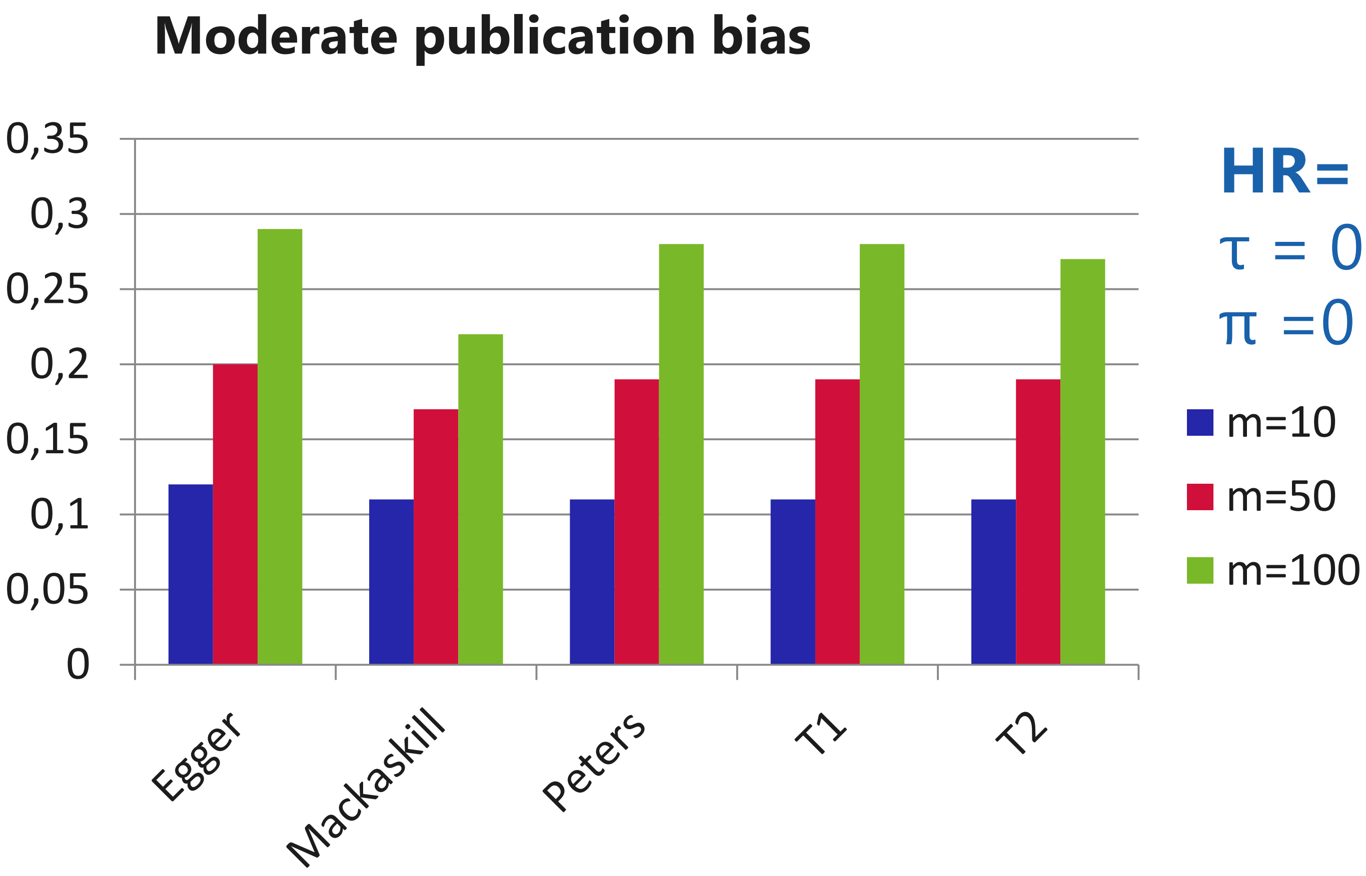
Simulation results: Type-I error

However, problems arise when HR<>1 !!



Other scenarios show similar problems for Eggers test and T2 when HR is extreme, m increases, and/or π approaches 0.

Simulation results: Power



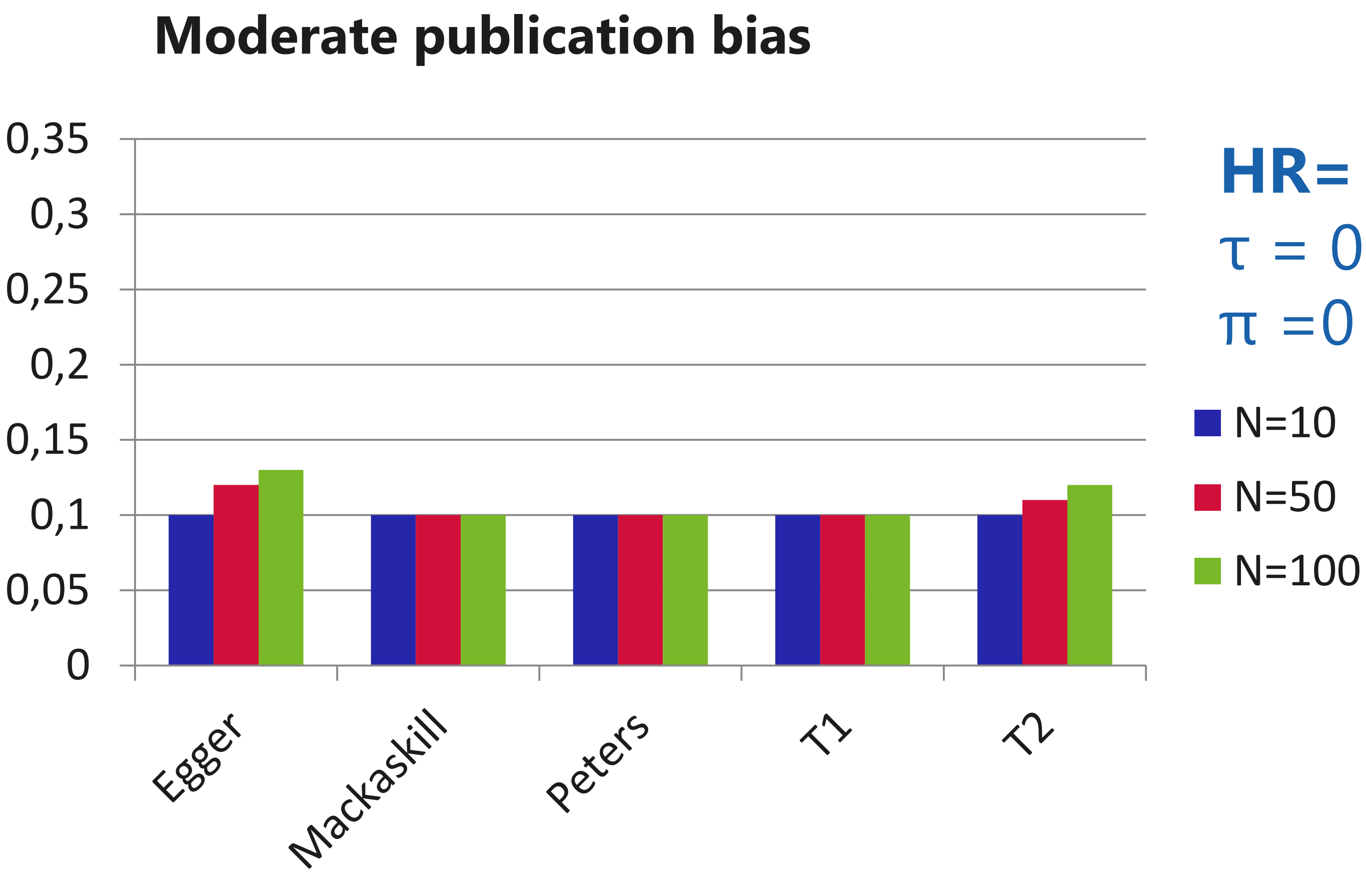
HR= 1

τ = 0

π =0

Note: T1 equals Peters test when π =0

Simulation results: Power



HR= 0.50

τ = 0

π =0

Note: T1 equals Peters test when π =0

Discussion

- Low power for **all tests**, unless many studies at hand
- Problematic type-I error for **Egger's test** and **T2**
- Testing presence publication bias much easier when true HR close to 1

	Type-I error	Power
Egger	- - -	++++
Macaskill	OK	+
Peters	OK	+++
T1	OK	+++
T2	-	++

(Preliminary) conclusions

- Use Peters or T1
- Further research needed on power in case of censoring!