A Meta-analytic Approach to Large Scale Competing Risks Regression

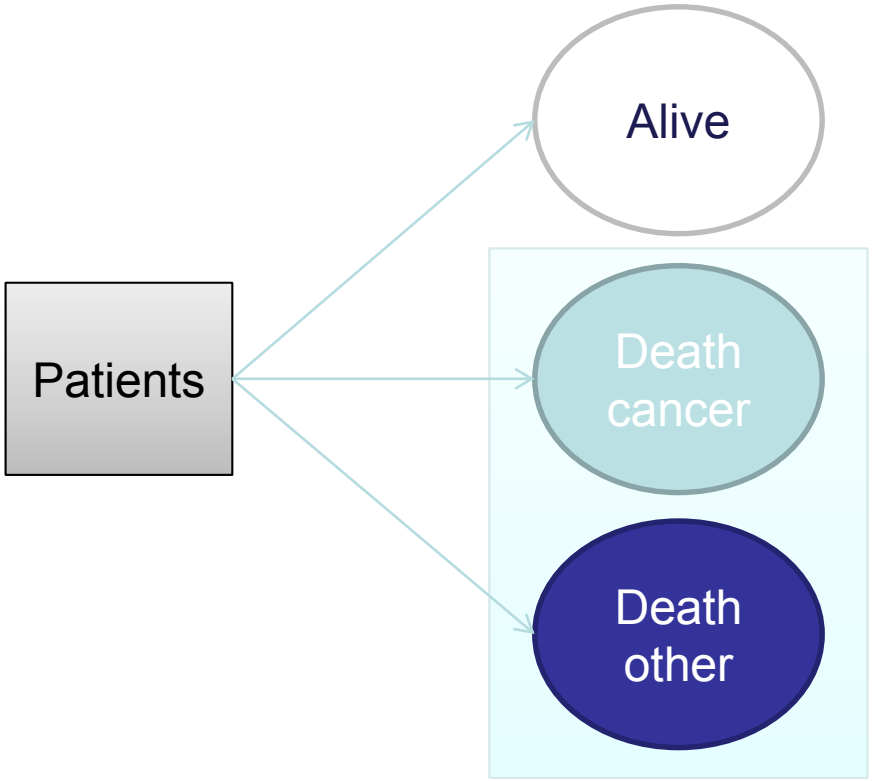
Chris Brown

CASI 2015, Cork

National Cancer Registry Ireland

Time-to-event question

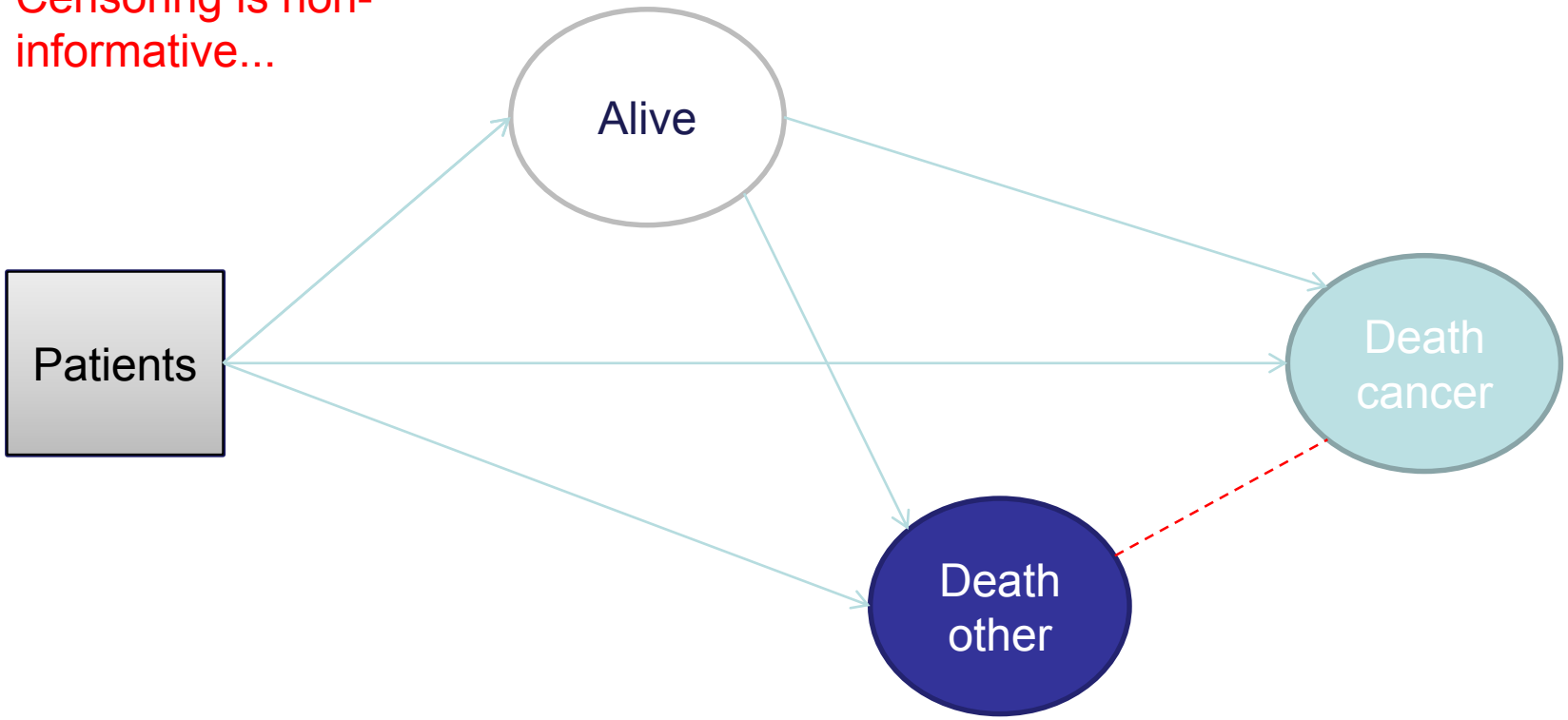
National Cancer Registry Ireland



Assumptions

National Cancer Registry Ireland

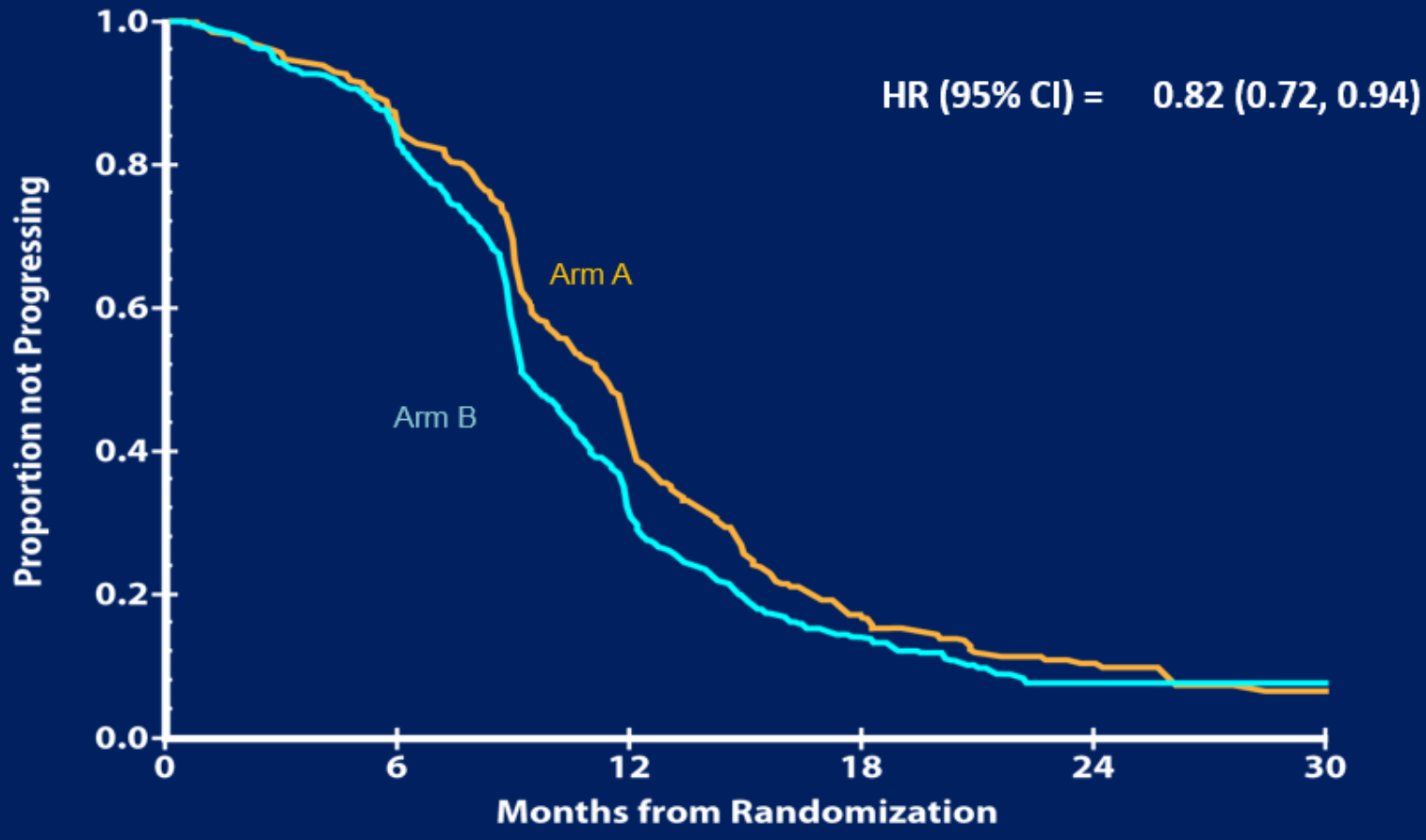
Censoring is non-informative...



Death due to other causes precludes cancer death...

Kaplan Meier / Cox regression

National Cancer Registry Ireland

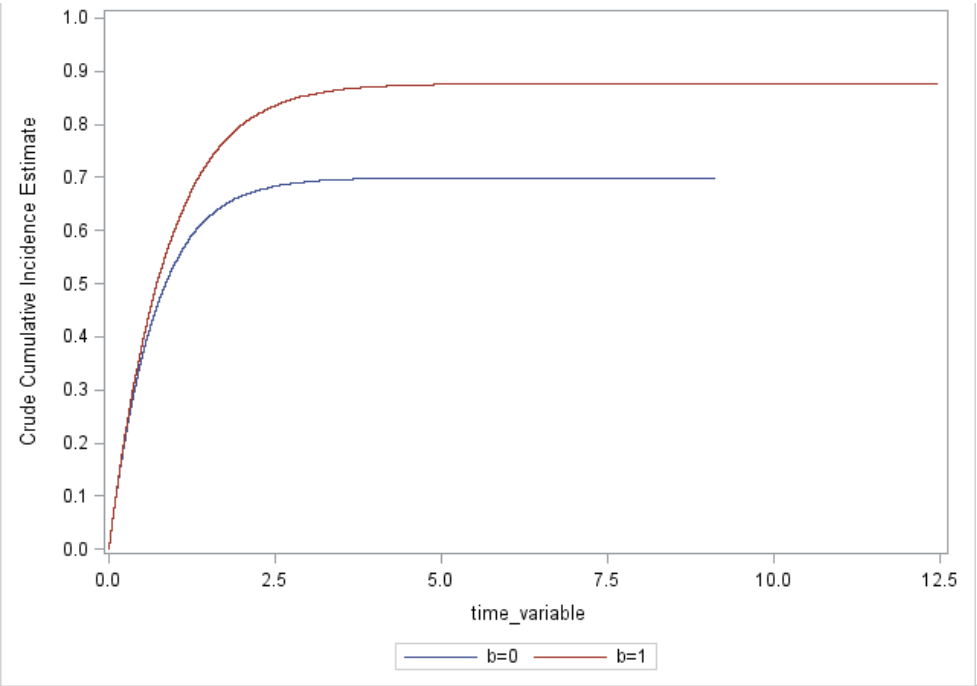


Pujade-Lauraine. Carboplatin & Pegylated Liposomal Doxorubicin versus Carboplatin & Paclitaxel in Relapsed, Platinum-sensitive Ovarian Cancer. ASCO 2009.

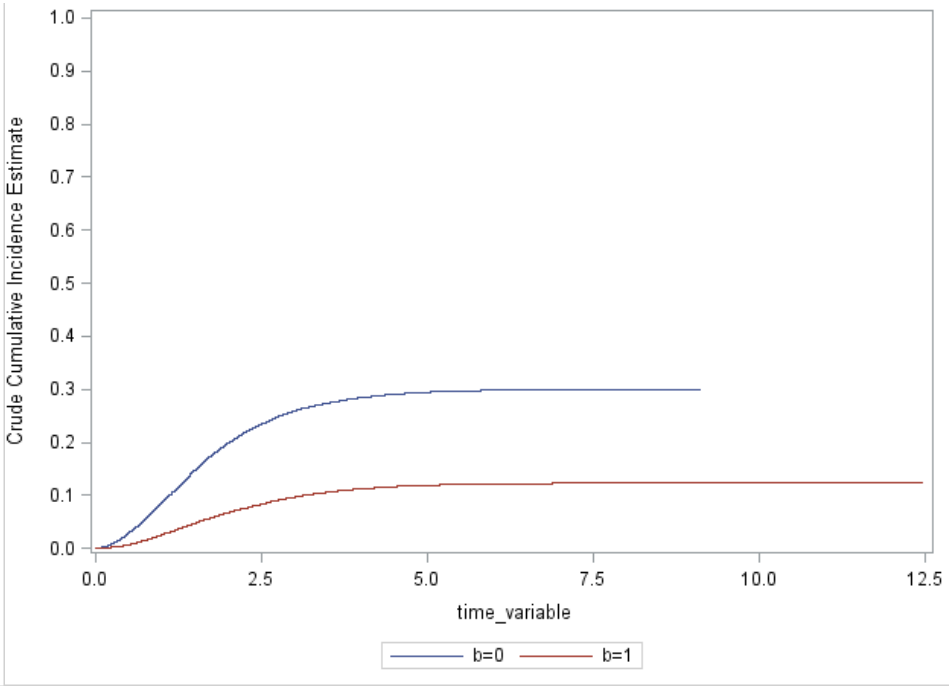
Cumulative Incidence

National Cancer Registry Ireland

Death cancer



Death other



Fine & Gray (FG) Regression

National Cancer Registry Ireland

Modelling effects on cumulative incidence.

- Weighted regression to account for the existence of competing risks

Implemented in all software packages

- e.g. SAS PSHREG

Computationally intensive with large datasets (or when large rates of competing risk)...

Kohl, Maria, Max Plischke, Karen Leondr, and Georg Heinze. (2014). PSHREG: A SAS Macro for Proportional and Nonproportional Subdistribution Hazards Re-gression. Computer Methods and Programs in Biomedicine, 118 (2)

Exponential computation

National Cancer Registry Ireland

Competing risks regression: Big Data issues from modest sample sizes

Katharine A. Kirby, University of California, San Francisco, California

Deborah E. Barnes, University of California, San Francisco, California

W. John Boscardin, University of California, San Francisco, California

Actual Sample Size	Approximate Number of Competing Events	Average Number of Pseudo Observations Per Competing Event	Fine-Gray Pseudo Sample Size	Run Time for PHREG	Run Time for PSHREG macro
1,000	300	250	77,000	<1 second	2 seconds
10,000	3,000	1,500	5M	<1 second	18 minutes
20,000	6,000	2,500	15M	<1 second	2 hours
40,000	12,000	4,000	46M	<1 second	10 hours
100,000	30,000	6,000	179M	1 second	72 hours
200,000	60,000	7,500	449M	2 seconds	200 hours

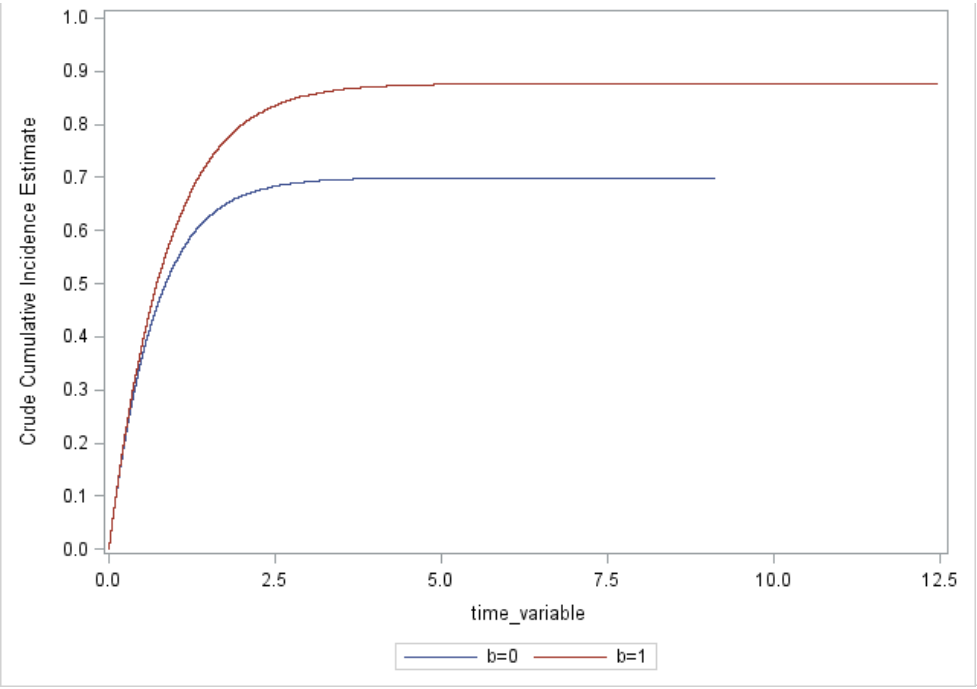
Table 1. Actual sample sizes, approximate number of competing events, average number of additional observations per competing event, effective sample sizes and run times for PROC PHREG and PSHREG macro (M=million).

http://www.wuss.org/Proceedings13/124_Paper.pdf

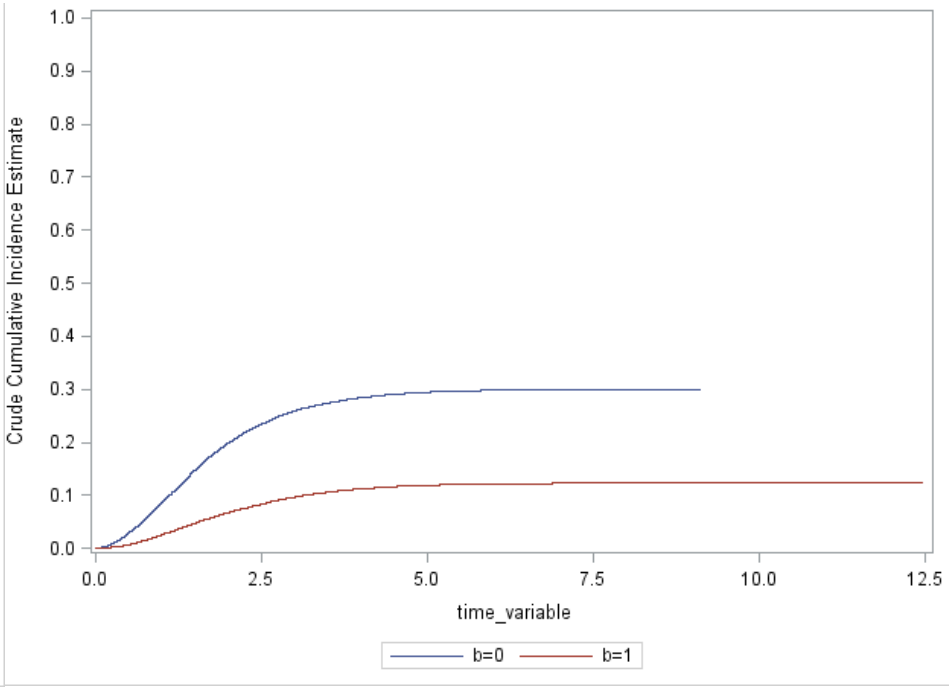
Simulation model

National Cancer Registry Ireland

Death cancer



Death other

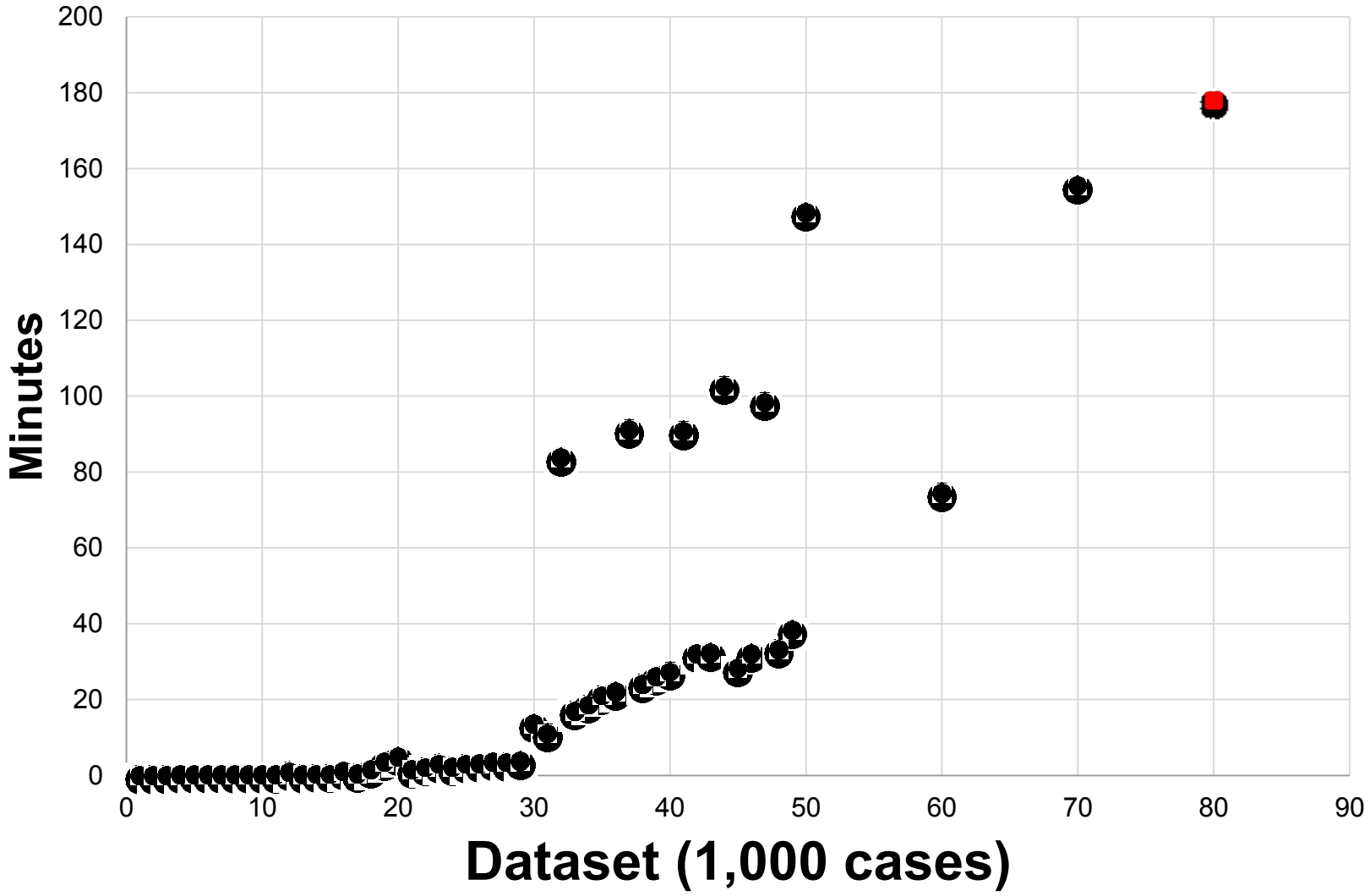


- Two event types, one covariate
- HR_{FG} death cancer = 2.745

Beyersmann, Jan, Aurlien Latouche, Anika Buchholz, and Martin Schumacher. (2009). Simulating Competing Risks Data in Survival Analysis. Statistics in Medicine, 28,(6)

Time to compute FG

National Cancer Registry Ireland



Meta-analysis solution

National Cancer Registry Ireland

1. Split large dataset into *i* (random) groups
2. Compute FG β estimates in each group
3. Combine using meta-analysis (IVW)

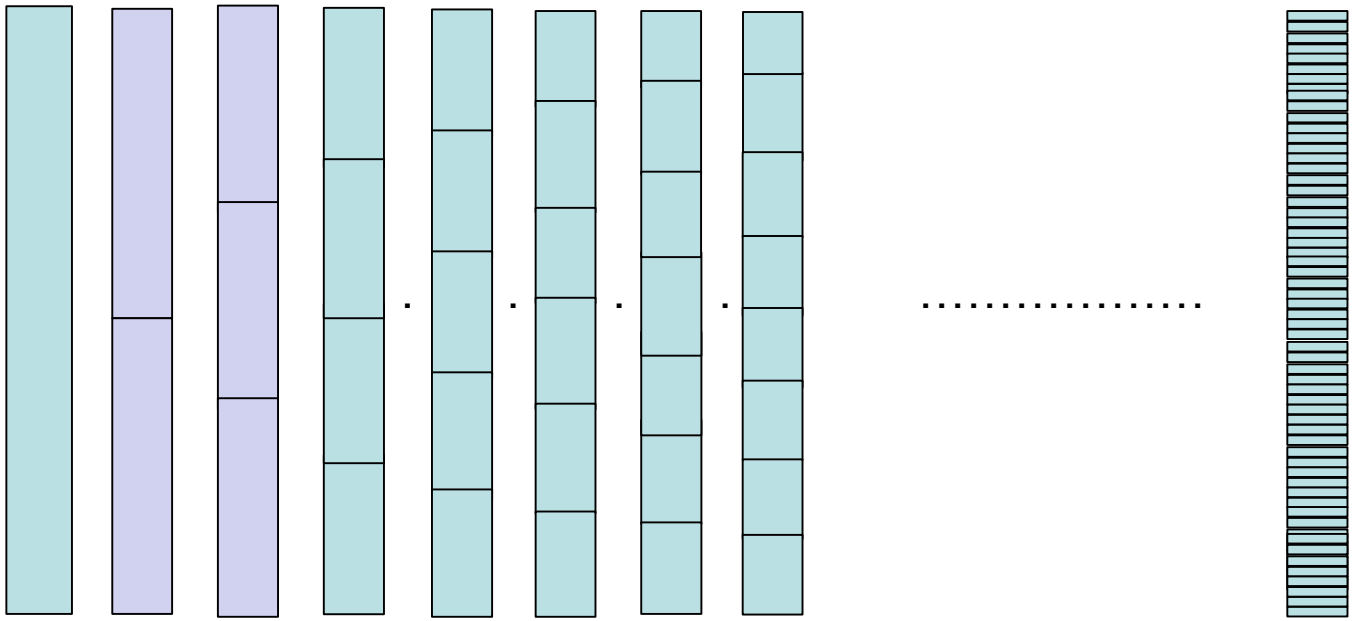
$$\hat{\beta} = \frac{\Sigma_i \beta_i / \sigma_i^2}{\Sigma_i 1 / \sigma_i^2}$$

KEY ASSUMPTION:

- Separate models estimating same quantity

Split the 80,000 cases

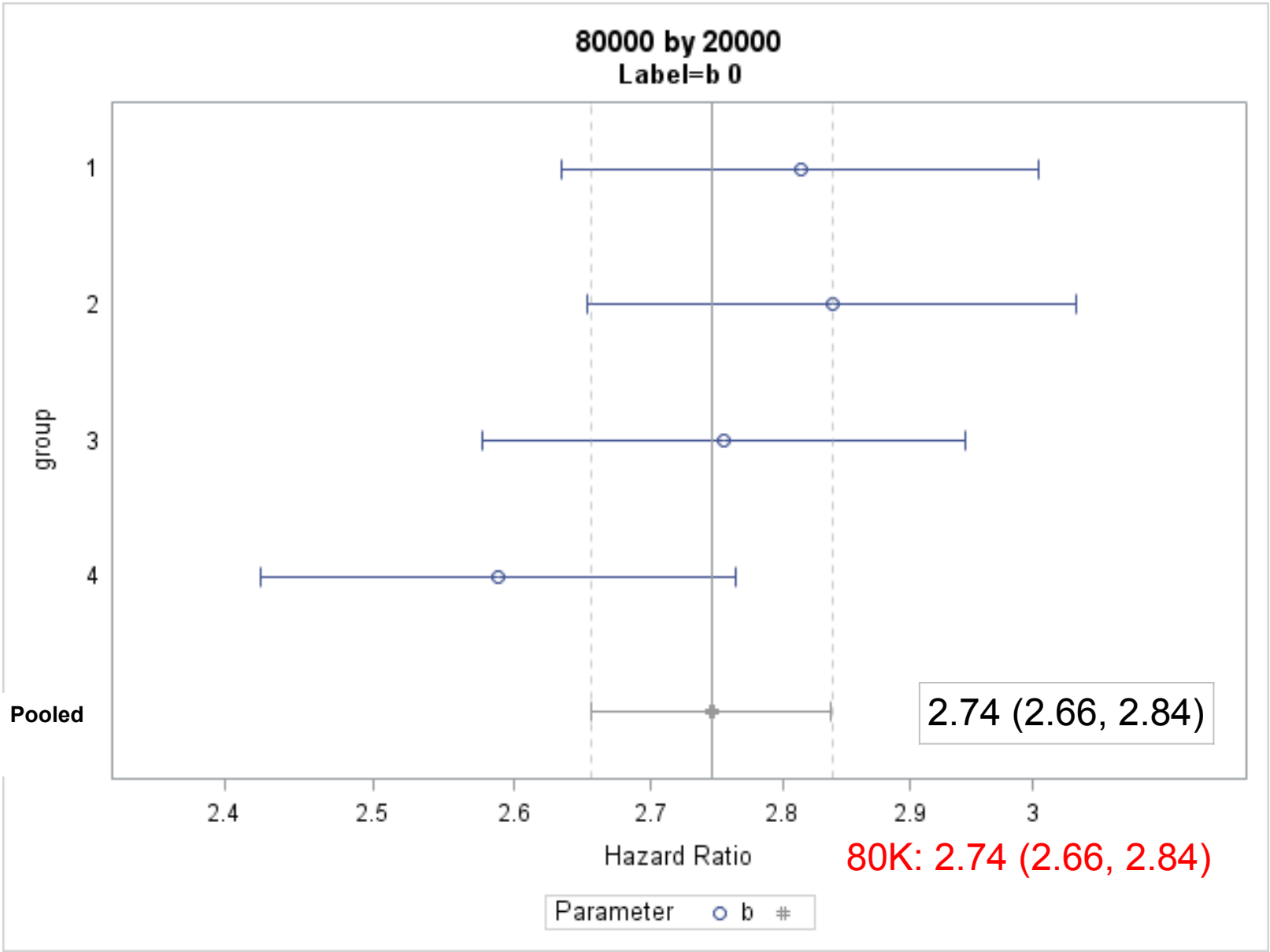
National Cancer Registry Ireland



Groups:	1	2	3	4 . 5 . 6 . 7 . 8	.....	80
Cases/group:	80,000	.....		10,000	.....	1,000

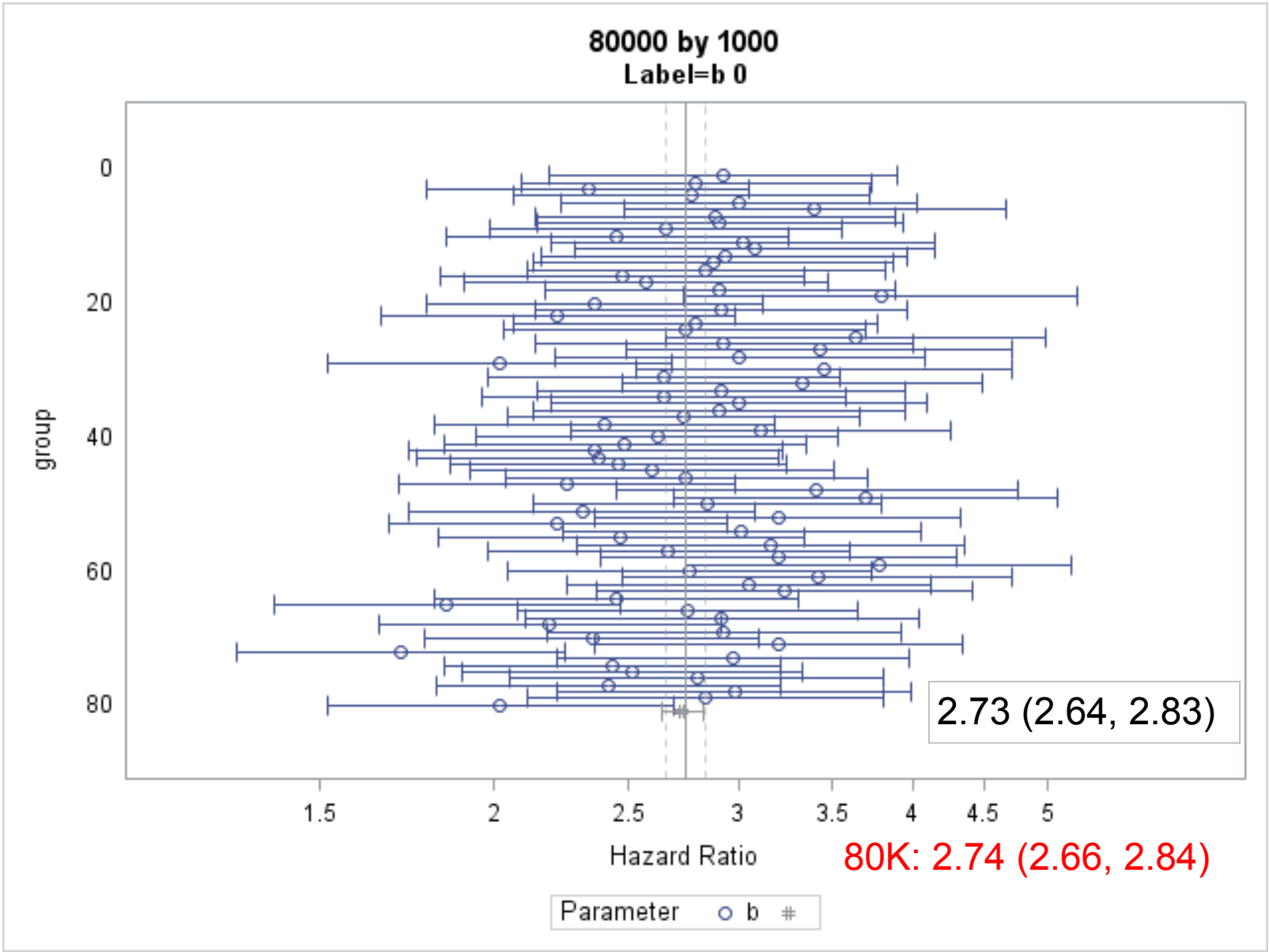
Groups of 20,000

National Cancer Registry Ireland



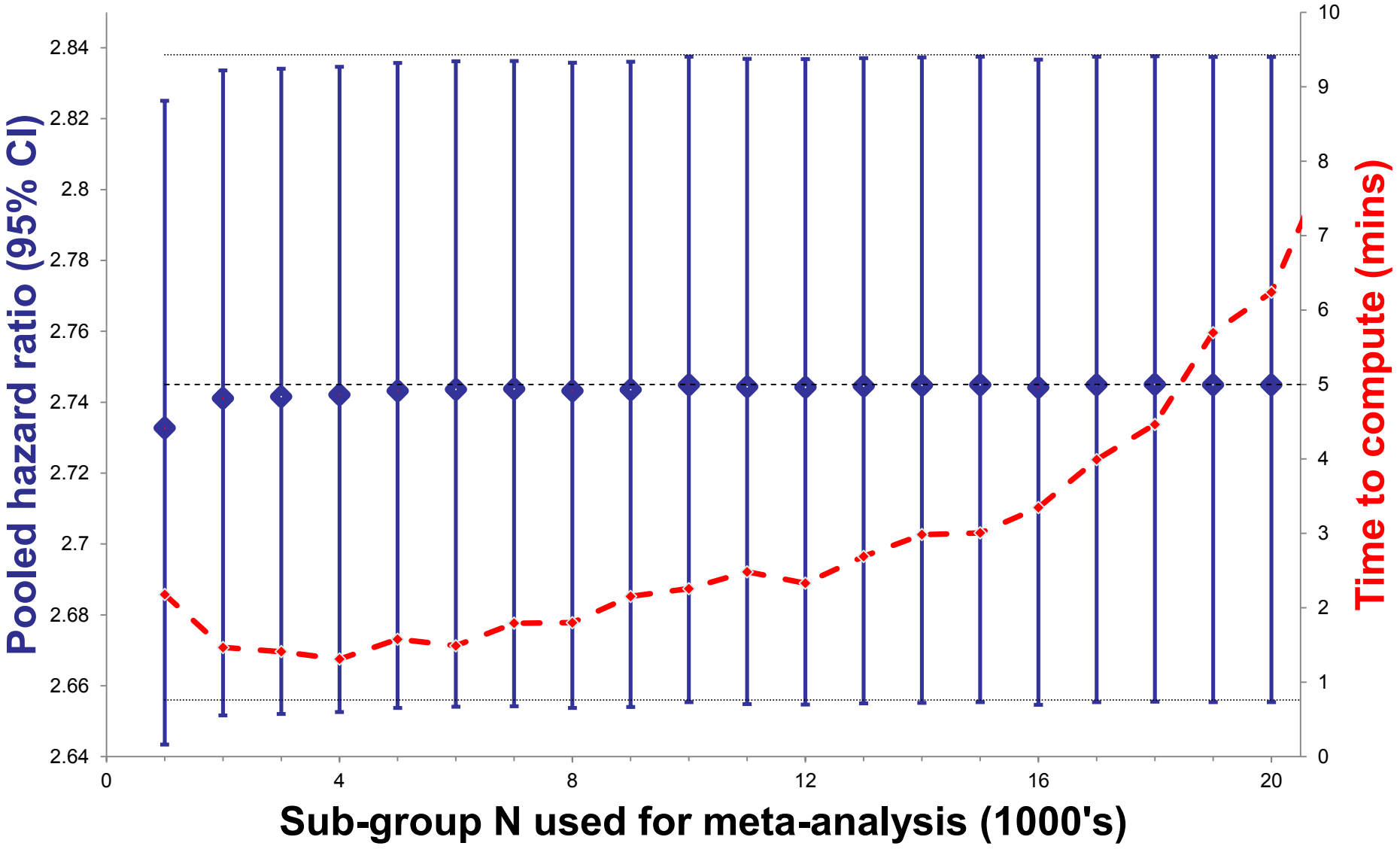
Groups of 1,000

National Cancer Registry Ireland



Total time vs. error

National Cancer Registry Ireland



Conclusion

National Cancer Registry Ireland

PROS

- Faster computationally (~135x)
- Little error introduced in estimates (<0.5%)

CONS

- Potential variability between groups (real data)
- Complexity of managing multiple covariates

Next steps

National Cancer Registry Ireland

- Illustrate reliability in:
  - Various competing risk scenarios
  - Real data from the Cancer Registry
- Models with multiple covariates
- R package for evaluating automatically

Thank you & Contact

National Cancer Registry Ireland

Questions or suggestions?

c.brown@ncri.ie

@CStatsAU

I'm supported by project funding from:

Health Research Board Ireland

HRB Health Research Board

The National Cancer Registry is funded by the Department of Health.

An Roinn Sláinte DEPARTMENT OF HEALTH