Efficiency of recurrent and time-to-first event methods in the presence of terminal events – Application to chronic heart failure trials

Patrick Schlömer
Bayer AG

Lifetime Data Science Conference
June 1st 2023, Raleigh, NC

Acknowledgements

- Philip Hougaard (Lundbeck)
- Henrik Ravn (Novo Nordisk)
- Arno Fritsch, Franco Mendolia (Bayer)
- Antje Jahn (Darmstadt University of Applied Sciences)
- Roger (London School of Hygiene and Tropical Medicine)
- Mouna Akacha, Frank Bretz, Tobias Mütze, Heinz Schmidli, Jiawei Wie (Novartis)

Heart Failure Trials

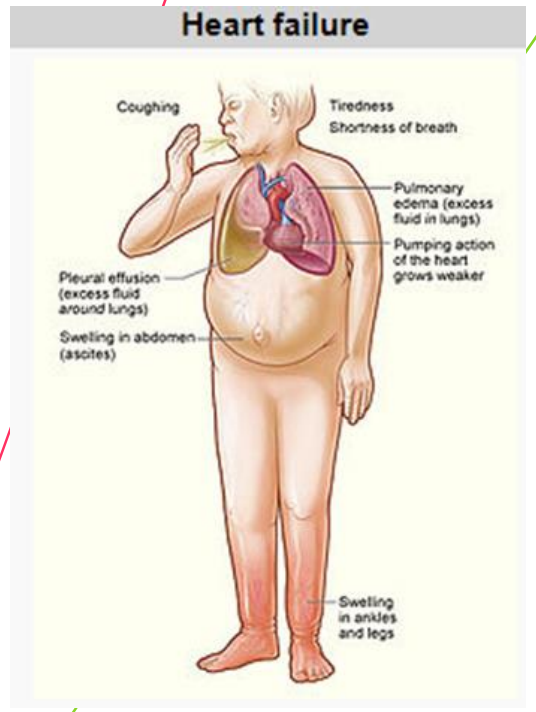
Motivation

Background: Heart Failure

Heart failure (HF) = Inability of the heart to pump sufficiently to maintain the blood flow that the body needs

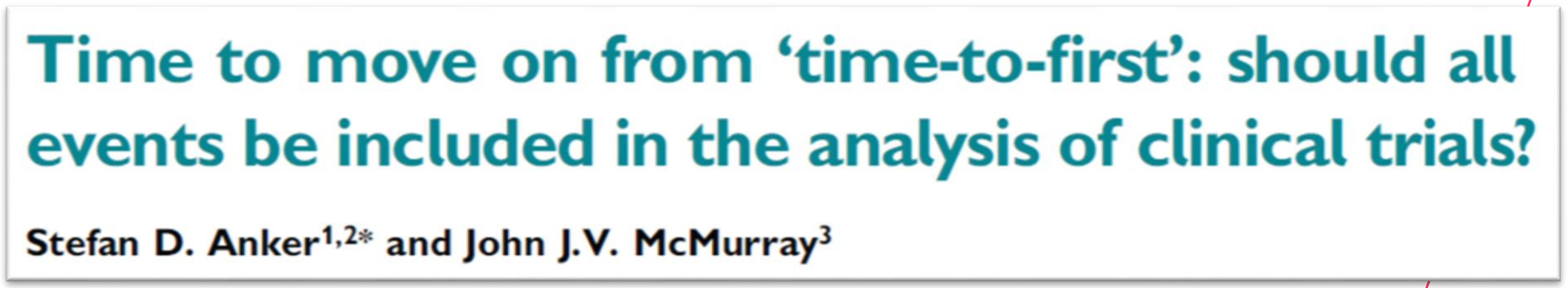
- **Major public health issue** with increasing prevalence
  - In 2020: 64.3 million people affected globally

- **Standard primary endpoint in heart failure trials**
  - Time-to-first hospitalization for heart failure (**HHF**) or cardiovascular death (**CVD**)



Source: Wikipedia

Primary Endpoint in Heart Failure Trials

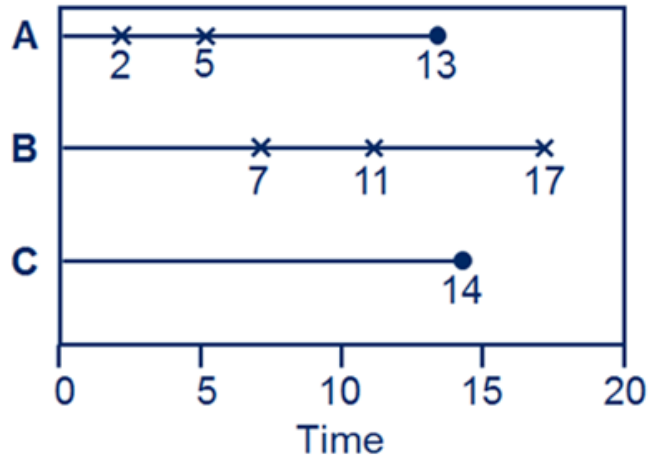

Time to move on from 'time-to-first': should all events be included in the analysis of clinical trials?

Stefan D. Anker[1,2*] and John J.V. McMurray[3]

- **Time-to-first** composite event analysis **ignores 40-50% of all CVDs and HHF events** (Anker & McMurray, 2012)
- Interested in **detailed investigation of potential power gains** when using recurrent event compared to time-to-first event endpoints in HF trials

Challenges with Recurrent Events in Heart Failure Trials

- **Challenge** of two types of events: HHFs & CVD
  - Risks **not independent**
  - CV death is **terminal**



- **Two endpoints** of potential **interest**
  - **HHF**: Focus on treatment effect on HHF (disregarding/adjusting for CVD)
  - **HHF+CVD**: Composite of HHF and CVD (CVD = last event)
  - Both endpoints evaluated with time-to-first event & different recurrent event methods

Methods

Methods Investigated (1)

- **Cox proportional hazards model** for time-to-first event
- **Quasi-Poisson model (QP)**
  - **Poisson model** with **overdispersion correction**
  - Investigated two estimation methods for **dispersion parameter**
- **Negative binomial regression (NB)**
  - **Poisson model** with **gamma-distributed** patient-specific **rates**
- **Lin-Wei-Yang-Ying (LWYY)**
  - **Andersen-Gill model** with **robust standard errors**
  - **Proportional rates model** with arbitrary baseline rate function

Methods Investigated (2)

- **Joint frailty model (JFM)** (Rogers et al, 2016)
  - **Jointly model** effect on **CVD and HHF** (linked by patient-specific **frailties terms z_i**)

$$\lambda_{\mathrm{CVD}} = \lambda_{0,\mathrm{CVD}}\, z_i^{\gamma}\, \exp(x_i\beta_2),$$
$$\lambda_{\mathrm{HHF}} = \lambda_{0,\mathrm{HHF}}\, z_i\, \exp(x_i\beta_1),$$

  - Distribution of z_i assumed to be log-normal (not gamma due to computational reasons)
  - Only applied to endpoint **HHF**

- **Win Ratio (WR)** (Pocock et al, 2012)
  - Based on unmatched pairwise comparison of patients between treatment groups
    1. Compare based on **time to CVD**
    2. If no winner or looser: Compare based on **time-to-first HHF**
  - Only considered for endpoint **HHF+CVD**

Simulation Study

Simulation Study – Set-Up

- **Joint frailty simulation** to capture dependency of **CVD & HHF** (Rogers et al, 2016)
- **Placebo rates (CVD & HHF)** based on **CHARM-preserved** and **TOPCAT** trials
- **Non-CVD** and **drug discontinuations** independently simulated
- **Treatment effects varied independently**
  - RR $(\exp(\beta_1))$ **Rate ratio for HHF**
  - HR $(\exp(\beta_2))$ **Hazard ratio for CVD**
- **90% power for LWYY** (two-sided) at 5% alpha level when HR=0.8 and RR=0.7 for endpoint **HHF+CVD**
  - **Sample size N=4350** (2175 patients per treatment arm)

Simulation Study – Type I Error

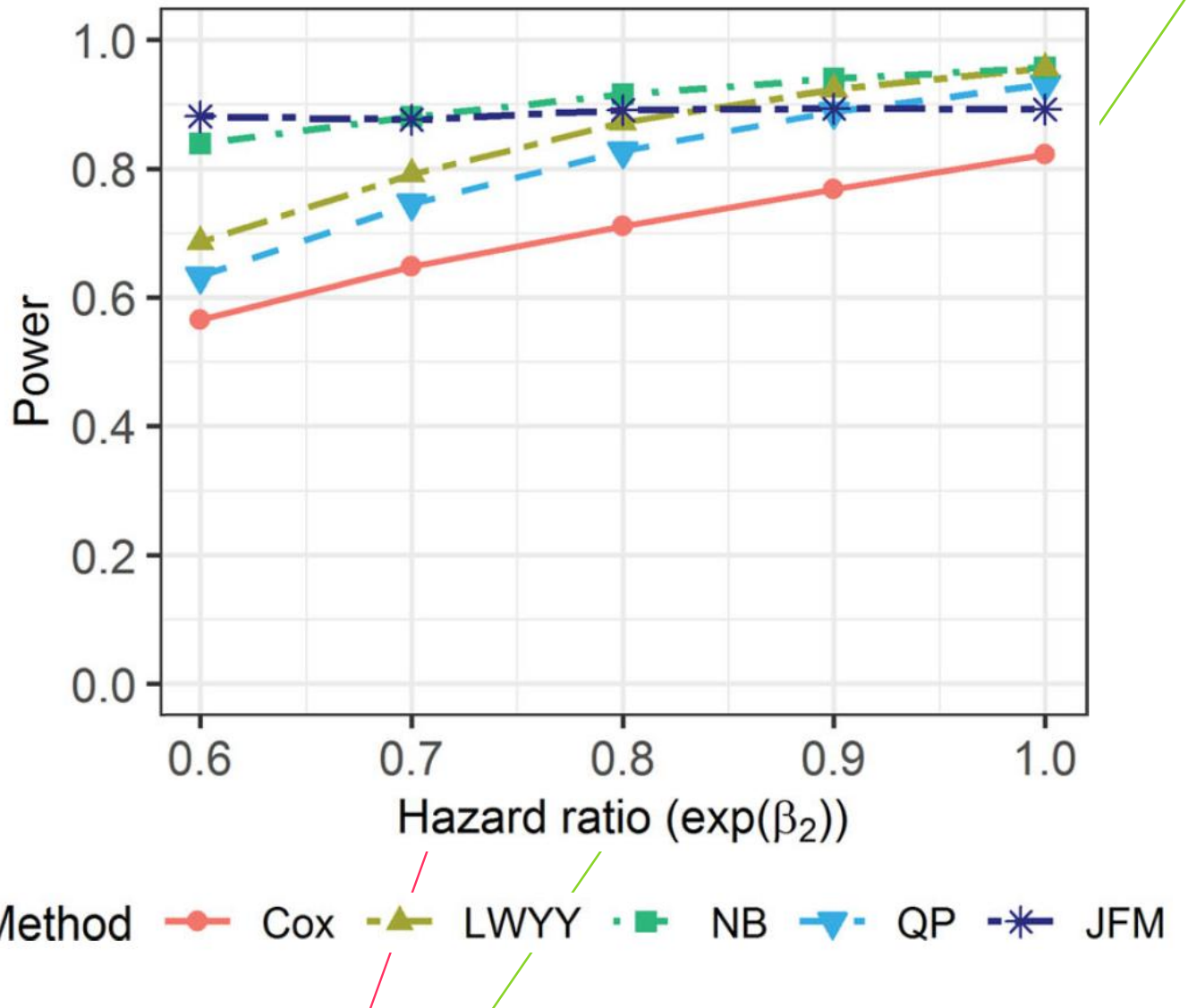
- Results shown based on simulating 10,000 studies

Table 2. Type I error rates (one-sided tests, nominal significance level $\alpha = 0.025$) for HHF and HHF+CVD with RR=1.

Endpoint	HR	LWYY	NB	Cox	JFM	Win ratio	Quasi-Poisson
HHF	0.6	0.000	0.004	0.005	0.024	–	0.000
	0.8	0.005	0.010	0.010	0.019	–	0.003
	1.0	0.023	0.025	0.023	0.028	–	0.015
	1.25	0.100	0.057	0.058	0.031	–	0.067
HHF+CVD	1.0	0.024	0.024	0.022	–	0.024	0.002
	1.25	0.023	0.004	0.004	–	0.002	0.001

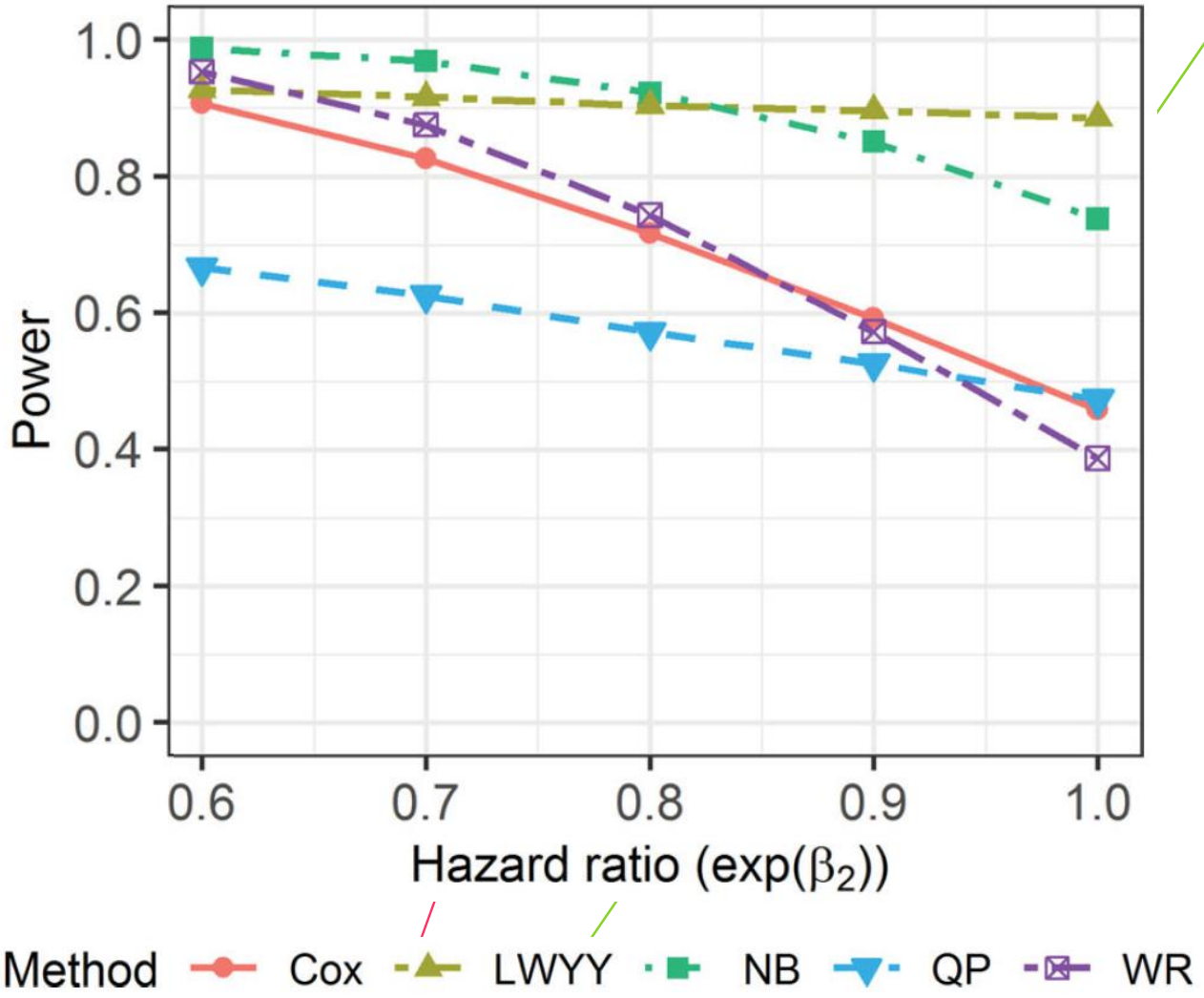
Simulation Study – Power HHF

- Fixed treatment effect on HHF (**RR=0.7**), **varying** effect on CVD (**HR**)
- **Substantial power gains** for recurrent event methods
- **Higher power** if **effect on CVD** is **lower**
  - **Advantage** for treatments with **worse effect on CVD** – undesirable
  - **JFM** seems able to **adjust for this**



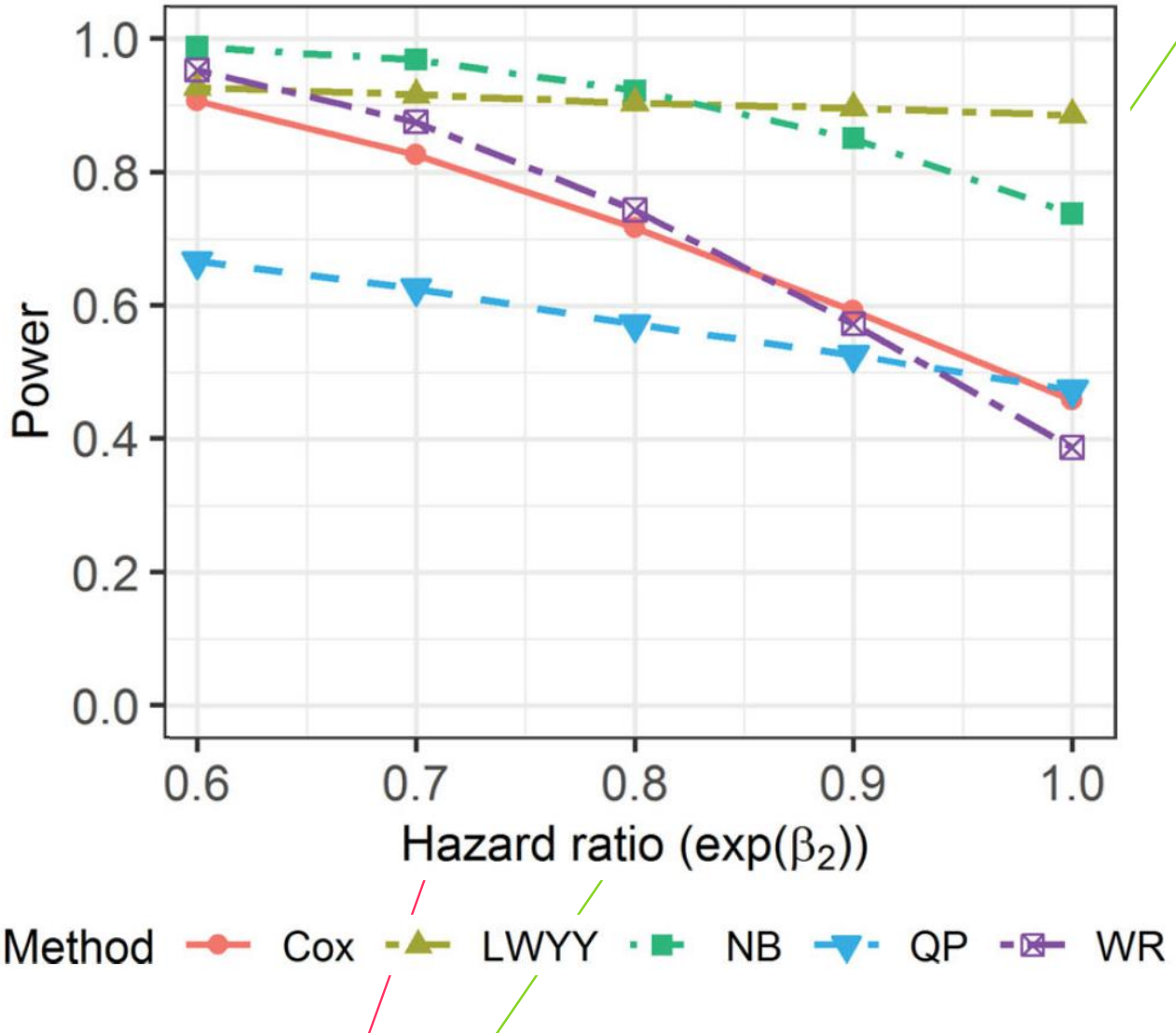
Simulation Study – Power HHF+CVD (1)

- **RR=0.7**, **varying HR**
- **Substantial power gains** for recurrent event methods (except QP)
- **Lower/similar power** if **CVD effect** is **lower**
  - **Counting additional CVD event counteracts** effect seen for HHF alone
  - Also seen in other simulation scenarios
- **WR** generally has **lower power** than recurrent event methods
  - **Higher power** than time-to-first event approach when **effect on HHF is small** and **effect on CVD is large**



Simulation Study – Power HHF+CVD (2)

- Investigated **low power of QP** further
  - Patients with an early event can lead to **overestimation of dispersion parameter**
  - Occurs for HHF+CVD when **patients die early**, less pronounced for HHF
  - **Alternative estimation method for dispersion parameter** by Fletcher (2012) instead of Pearson statistic overcorrected and **inflates type I error**



*Bootstrap-Based Efficiency Comparison**

*JFM not included due to computational burden

Val-HeFT trial

A Randomized Trial of the Angiotensin-Receptor Blocker Valsartan in Chronic Heart Failure

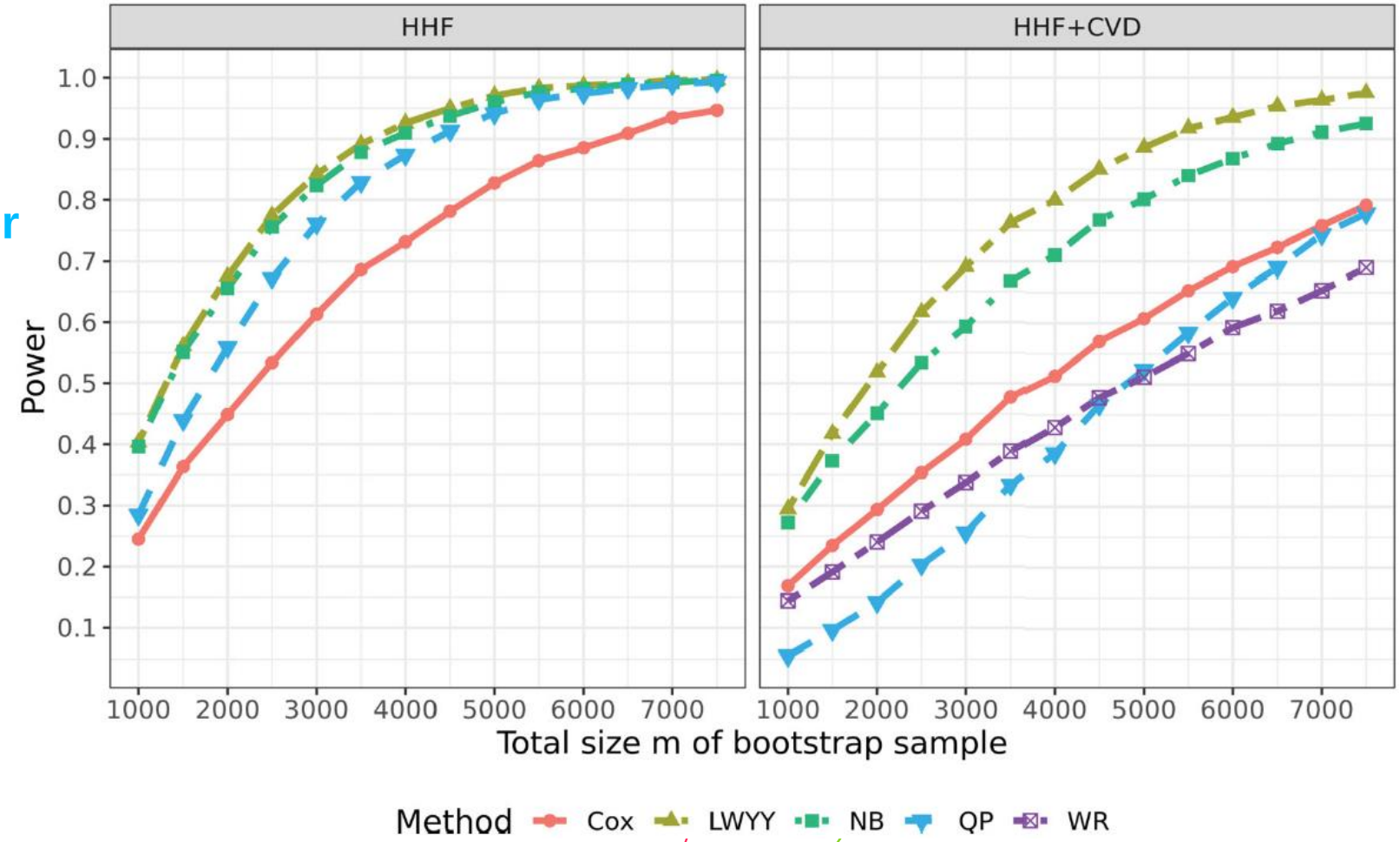
Jay N. Cohn, M.D., and Gianni Tognoni, M.D. for the Valsartan Heart Failure Trial Investigators*

- 5010 patients with chronic HF, randomly assigned to valsartan or placebo (1:1)
- **No effect on mortality** (HR=1.02; 98%-CI (0.88-1.18))
- Rate of **first occurrence of composite endpoint*** reduced by 13 % for valsartan
  - Mainly **driven by effect on HHF** (HR=0.84; 95%-CI (0.75-0.95))

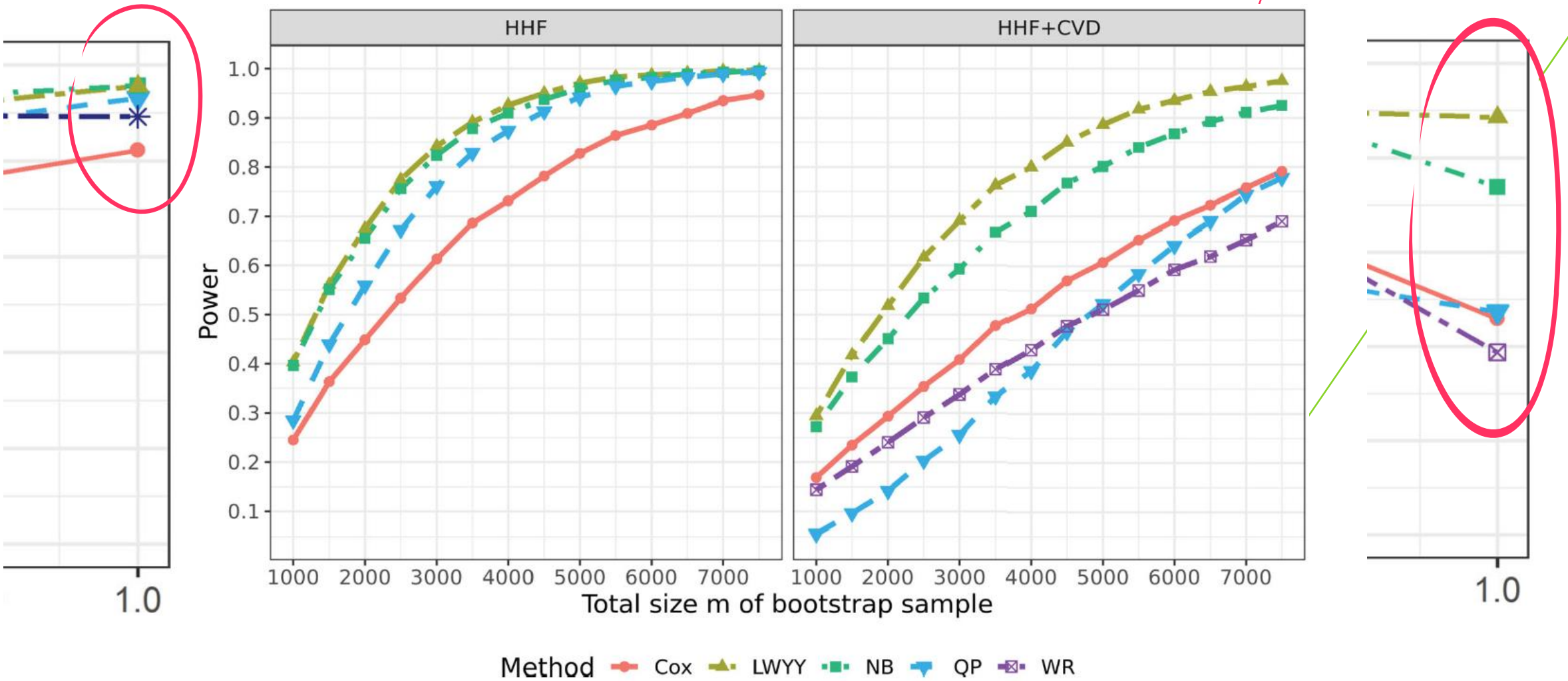
*Composite endpoint included: Death from any cause, HHF, Cardiac arrest with resuscitation, Intravenous therapy

Val-HeFT trial – bootstrap results

- **HHF**
  - **LWYY and NB** have roughly **same power**
  - **Cox and QP** have **lower power**
- **HHF+CVD**
  - Substantial **power gain for LWYY and NB** compared to time-to-first event analysis
  - **LWYY > NB**
  - **QP and Win Ratio** have **lower power** than time-to-first event analysis



ValHeFT results **in line with simulation results** for scenarios with **no treatment effect on CVD**



PARADIGM-HF trial

Angiotensin–Neprilysin Inhibition versus Enalapril in Heart Failure

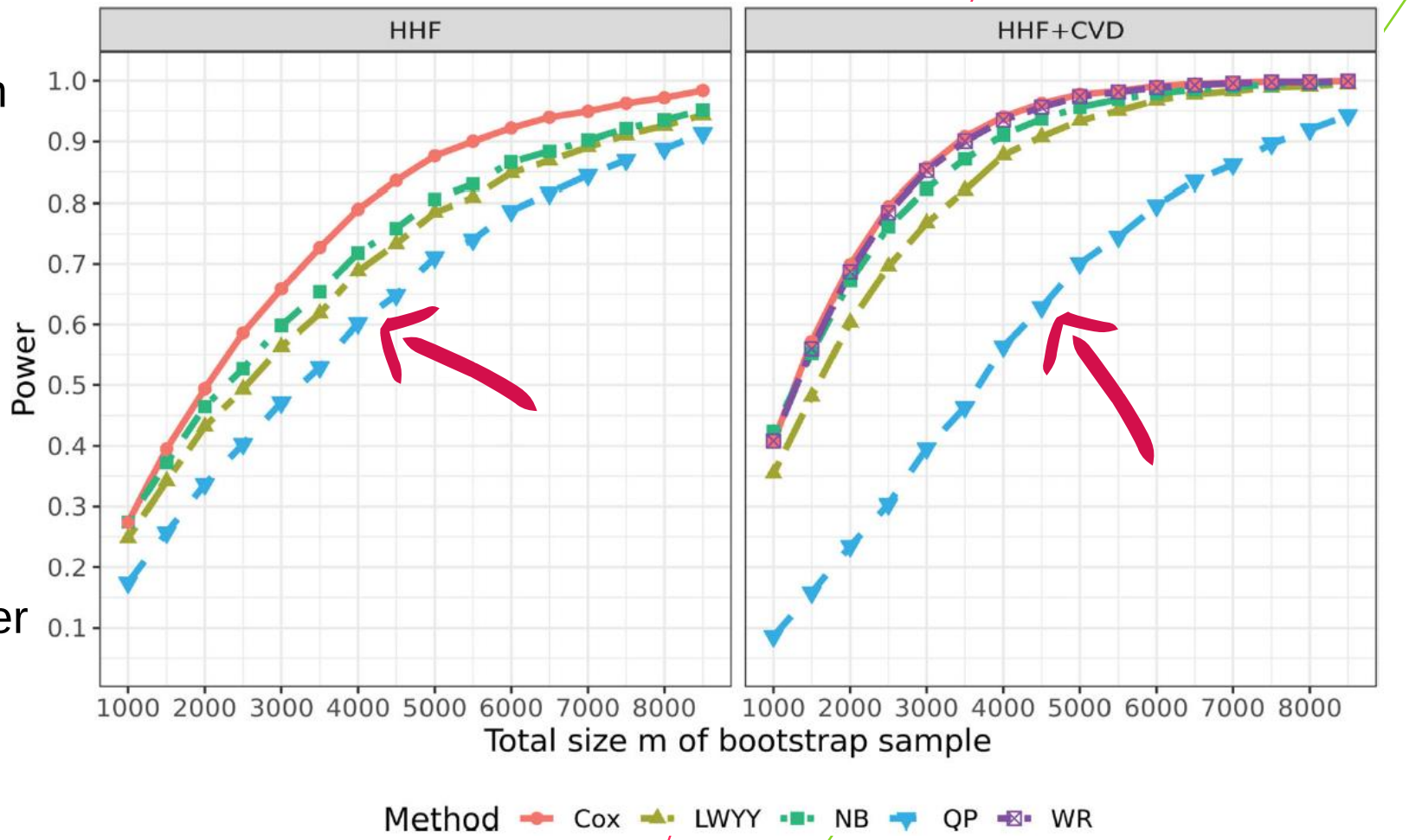
John J.V. McMurray, M.D., Milton Packer, M.D., Akshay S. Desai, M.D., M.P.H., Jianjian Gong, Ph.D., Martin P. Lefkowitz, M.D., Adel R. Rizkala, Pharm.D., Jean L. Rouleau, M.D., Victor C. Shi, M.D., Scott D. Solomon, M.D., Karl Swedberg, M.D., Ph.D., and Michael R. Zile, M.D. for the PARADIGM-HF Investigators and Committees*

- 8442 patients with chronic HF*, randomly assigned to sacubitril-valsartan or enalapril (1:1)
- Rate of **first occurrence of composite endpoint** (CVD+HHF) reduced by 20 % for sacubitril-valsartan compared to enalapril
  - **Effect identical for components**
    (CVD: HR=0.80; 95%-CI (0.71-0.89), HHF: HR=0.79; 95%-CI (0.71-0.89))

*chronic HF with reduced left ventricular ejection fraction

PARADIGM-HF trial – bootstrap results

- **Cox and WR** have **higher power** than recurrent event methods (both **HHF** & **HHF+CVD**)
  - Possible explanations
    - **High drug discontinuation rate** after first event (40 %)
    - Positive **treatment effect on CVD** (similar magnitude as HHF effect)
- Clagget et al. (2018) reported power gains of recurrent event methods if **drug disc. after HHF ≤ 30%**
- **Low power for QP** aligned with results of simulation study



Summary and Discussion

Summary and discussion

- **Higher power for recurrent event methods** in many situations for **HHF and HHF+CVD**
  - **Exception**: Many **drug disc.** after HHF; large **CVD effect**
  - Issue for QP in case of early events
- **Undesirable behavior for HHF**
  - Power increase with smaller effect on CVD due to **selection bias**
  - Inflated Type I error for **detrimental CVD effect**
  - Methods **not recommended for HHF** unless **no effect on CVD** can be assumed
  - **Exception: JFM**
- **HHF+CVD**
  - **Including CVD as event** seems to **prevent issues** described for HHF only
  - At least for scenarios that are realistic for HF trial

14 April 2020
EMA/CHMP/SAWP/120610/2020
Committee for Medicinal Products for Human Use (CHMP)

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Qualification opinion of clinically interpretable treatment effect measures based on recurrent event endpoints that allow for efficient statistical analyses

Efficiency Comparison of Analysis Methods for Recurrent Event and Time-to-First Event Endpoints in the Presence of Terminal Events—Application to Clinical Trials in Chronic Heart Failure

Arno Fritsch[a], Patrick Schlömer[b], Franco Mendolia[a], Tobias Mütze[c], and Antje Jahn-Eimermacher[d] on behalf of the Recurrent Event Qualification Opinion Consortium*

[a]Bayer AG, Pharmaceuticals, Wuppertal, Germany; [b]Bayer AG, Pharmaceuticals, Berlin, Germany; [c]Novartis Pharma AG, Basel, Switzerland; [d]Department of Mathematics and Natural Sciences, Darmstadt University of Applied Sciences, Darmstadt, Germany

Estimands for Recurrent Event Endpoints in the Presence of a Terminal Event

Heinz Schmidli[a], James H. Roger[b], and Mouna Akacha[c]

[a]Statistical Methodology, Novartis, Basel, Switzerland; [b]Medical Statistics Department, London School of Hygiene & Tropical Medicine, London, UK; [c]Statistical Methodology, Novartis, Basel, Switzerland, on behalf of the Recurrent Event Qualification Opinion Consortium*

Properties of Two While-Alive Estimands for Recurrent Events and Their Potential Estimators

Jiawei Wei[a], Tobias Mütze[b], Antje Jahn-Eimermacher[c], and James Roger[d]

[a]Novartis Institutes for Biomedical Research Co., Shanghai, China; [b]Novartis Pharma AG, Basel, Switzerland; [c]Darmstadt University of Applied Sciences, Darmstadt, Germany; [d]London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom on behalf of the Recurrent Event Qualification Opinion Consortium*

Thank you!

patrick.schloemer@bayer.com

References

- **Anker, S. D. and McMurray, J. J. (2012).** Time to move on from “time-to-first”: should all events be included in the analysis of clinical trials? Eur. Heart J. 33, 2764-2765.
- **Cohn, J., Tognoni, G. and the Valsartan Heart Failure Trial Investigators (2001).** A Randomized Trial of the Angiotensin-Receptor Blocker Valsartan in Chronic Heart Failure. New England Journal of Medicine, 345, 1667–1675.
- **Fletcher, D. (2012).** Estimating Overdispersion When Fitting a Generalized Linear Model to Sparse Data. Biometrika, 99, 230–237.
- **Fritsch, A., Schlömer, P., Mendolia, F., Mütze, T., Jahn-Eimermacher, A. (2023).** Efficiency Comparisn of Analysis Methods for Recurrent Event and Time-to-First Event Endpoints in the Presence of Terminal Events – Application to Clinical Trials in Chronic Heart Failure. Stat. in Biopharm. Research, 15, 268-279.
- **McMurray, J. J., Packer, M., Desai, A. S., Gong, J., Lefkowitz,M. P., Rizkala, A. R., Rouleau, J. L., Shi, V. C., Solomon, S. D., Swedberg, K. and Zile, M. R. for the PARADIGM-HF Investigators and Committees. (2014).** Angiotensin–Neprilysin Inhibition Versus Enalapril in Heart Failure, New England Journal of Medicine. 371, 993–1004.
- **Pitt, B., Pfeffer, M. A., Assmann, S. F., Boineau, R., Anand, I. S., Clagett, B., Clausell, N., Desai, A. S., Diaz, R., Fleg, J. L., Gordeev, I., Harty, B., Heitner, J. F., Kenwood, C. T., Lewis, E. F., O’Meara, E., Probstfield, J. L., Shaburishishvili, T., Shah, S. J., Solomon, S. D., Sweitzer, N.K., Yang, S. and McKinlay, S. (2014).** Spironolactone for heart failure with preserved ejection fraction. New Engl. J. of Med. 370, 1383-1392
- **Pocock, S. J., Ariti, C. A., Collier, T. J. and Wang, D. (2012).** The win ratio: a new approach to the analysis of composite endpoins in clinical trials based on clinical priorities. Eur. J. Heart Failure, 33, 176-182.
- **Rogers, J. K., Yaroshinsky, A., Pocock, S. J., Stokar, D. and Pogoda, J. (2016).** Analysis of recurrent events with associated informative dropout time: Application of the joint frailty model. Stats in Med., 35, 2195-2205.
- **Schmidli, H., Roger, J. H. and Akacha, M. (2023).** Estimands for Recurrent Event Endpoints in the Presence of a Terminal Event. Stat. in Biopharm. Research, 15, 238-248.
- **Yusuf, S., Pfeffer, M. A., Swedberg, K., Granger, C. B., Held, P., McMurray, J. J. V., Michelson, E. L., Olofsson, B. and Östergren, J. (2003).** Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved trial. Lancet, 362, 777-781
- **Wei, J., Mütze, T., Jahn-Eimermacher, A. and Roger J. H. (2023).** Properties of Two While-Alive Estimands for Recurrent Events and Their Potential Estimators. Stat. in Biopharm. Research, 15, 257-267.

Back-up

Simulation Study – Set-Up (1)

- Model situations **as observed in previous HF trials**
- **Gap times** between HHFs and **CVD & HHF** processes **strongly correlated**
  - **Joint frailty model** to incorporate dependency (gamma-distributed frailties)
  - Exponent for CVD frailty term $\gamma = 0.75$, similar to previous trials (Rogers et al, 2016)
    - Lower variability for CVD than HHF
- **Annualized placebo rate of CVD** 4.0 (events / 100 patient-years)
  - CHARM-Preserved 3.9; TOPCAT BNP Stratum 3.9
- **Annualized placebo rate of first composite event** 9.0
  - CHARM-Preserved 9.1; TOPCAT BNP Stratum 8.5
- **Non-CVD** independently simulated to be around **30% of total deaths**

Simulation Study – Set-Up (2)

- Observed **ratio recurrent to first composite events** = 1.8 (Anker and McMurray, 2012)
  - Chose frailty variance ϕ so that this ratio is observed
- **Study duration** 5 years / **Patient recruitment** 3 years
- It is assumed that 5% of patients **discontinue treatment** each year
- **Treatment effects varied independently**
  - RR $(\exp(\beta_1))$ **Rate ratio for HHF**
  - HR $(\exp(\beta_2))$ **Hazard ratio for CVD**
- **90% power for LWYY** (two-sided) at 5% alpha level when HR=0.8 and RR=0.7 for endpoint **HHF+CVD**
  - **Sample size N=4350** (2175 patients per treatment arm)

Bootstrap-Based Efficiency Comparison using Clinical Trial Data

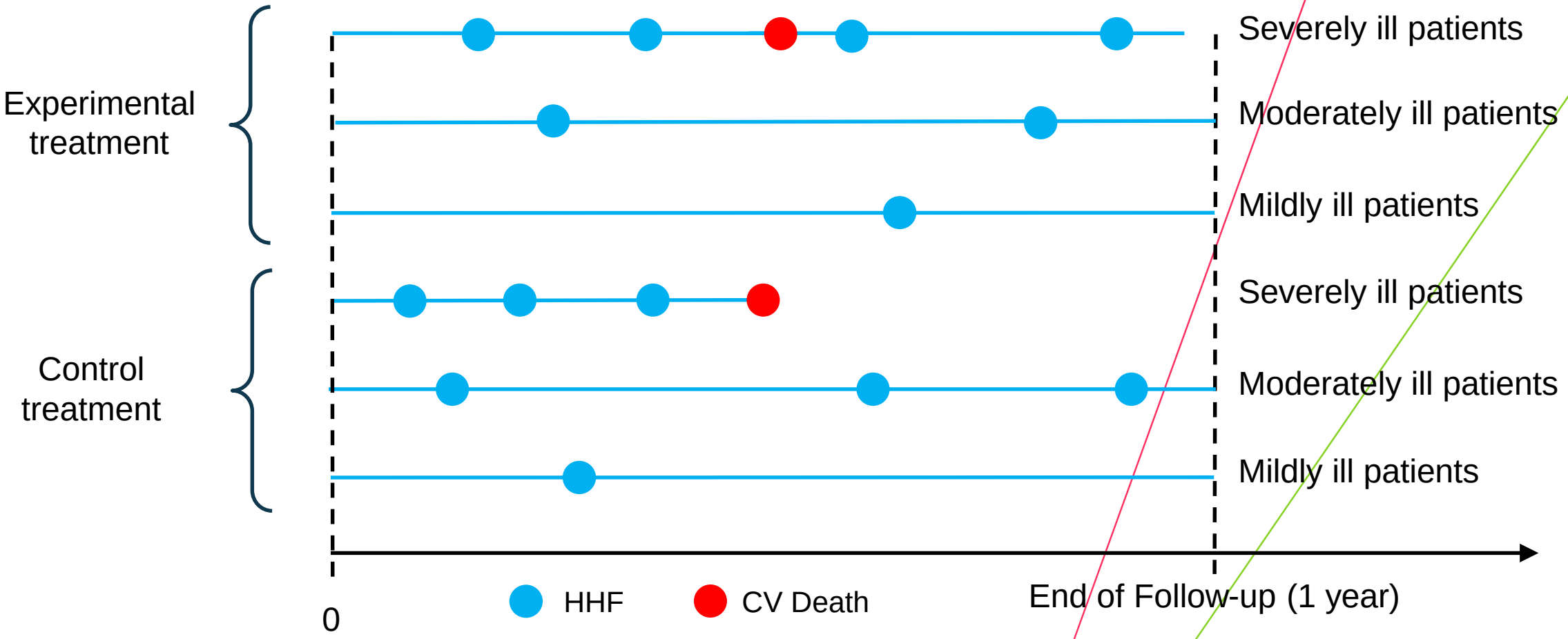
- **Resampling of clinical trial data** to closer **capture clinical trial setting**

- **Bootstrap approach** for given sample size m:

  1. For b = 1, . . . , B, with B = 500,000, repeat steps a) and b)
     a) Generate data for treatment (control) group by **drawing with replacement** a sample of size m/2 from the treatment (control) group of the clinical trial data.
     b) Analyze the bootstrap sample using the analysis method of interest. Store the **one-sided p-value p_b**.
  2. The power of the analysis method is given by $\mathbf{Power} = \sum_{b=1}^{B} I\{p_b \leq 0.025\}/B$

- JFM is not used, due to the computational burden

Why advantage for treatment with worse CV death effect?



Exact Values for Data-Generating Model Parameters

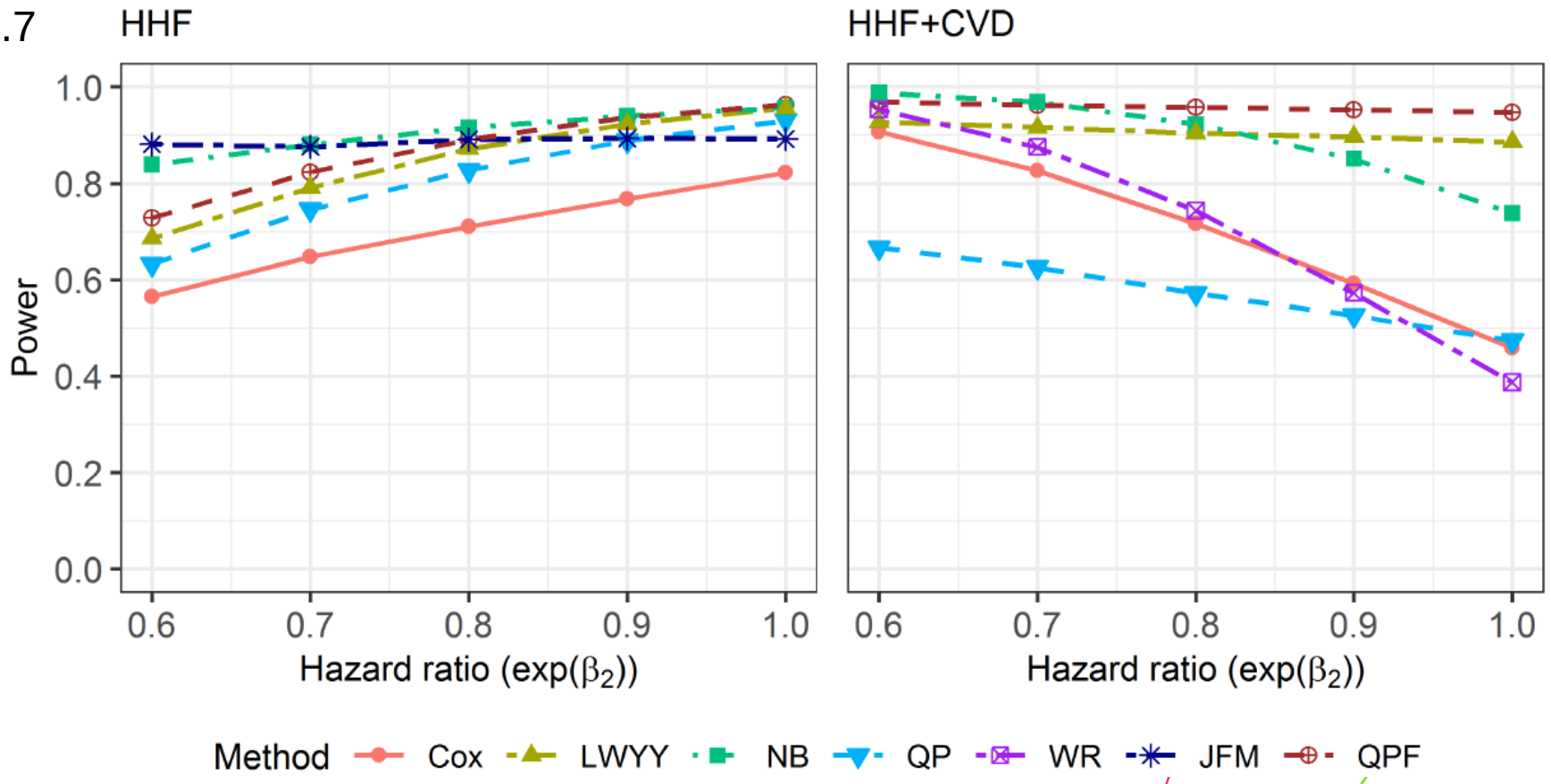
	$\lambda_{0,CVD}$	$\lambda_{0,HHF}$	ϕ	γ	λ_{nonCVD}	$\lambda_{trtDisc}$
Base case	0.07032	0.15444	5.7	0.75	0.01716	0.05129
20% Drug Disc. after HHF	0.07032	0.15444	5.7	0.75	0.01716	–
40% CV mortality	0.77400	0.15444	5.7	0.75	0.07920	0.05129

Trial results for endpoint HHF+CVD

Observed treatment effects RR/HR (with 95% CI) and p-values				
	LWYY	NB	Cox	Win Ratio
ValHeFT	0.83 (0.75 – 0.93) p=0.002	0.83 (0.72 – 0.95) p=0.008	0.89 (0.81 – 0.98) p=0.020	1.13 (1.03 – 1.24) p=0.010
CHARM-Preserved	0.78 (0.65 – 0.93) p=0.006	0.75 (0.62 – 0.91) p=0.003	0.86 (0.74 – 1.00) p=0.050	1.17 (0.99 – 1.39) p=0.065
CHARM-Added	–	0.75 (0.62 – 0.91) p=0.003	0.83 (0.74 – 0.94) p=0.003	1.30 (1.13 – 1.50) p<0.001
CHARM-Alternative	–	0.65 (0.51 – 0.82) p<0.001	0.77 (0.67 – 0.89) p<0.001	1.42 (1.20 – 1.70) p<0.001
PARADIGM-HF	0.79 (0.71 – 0.87) P<0.001	0.76 (0.67 – 0.85) P<0.001	0.80 (0.73 – 0.87) P<0.001	–
PARAGON-HF	0.87 (0.75 – 1.01) p=0.059	0.87 (0.74 – 1.01) p=0.065	0.92 (0.814 – 1.03) p=0.154	–

Simulation Study – Quasi-Poisson with Fletcher‘s Method (1)

RR=0.7



// High power for QP with Fletcher‘s method (QPF) due to type I error inflation

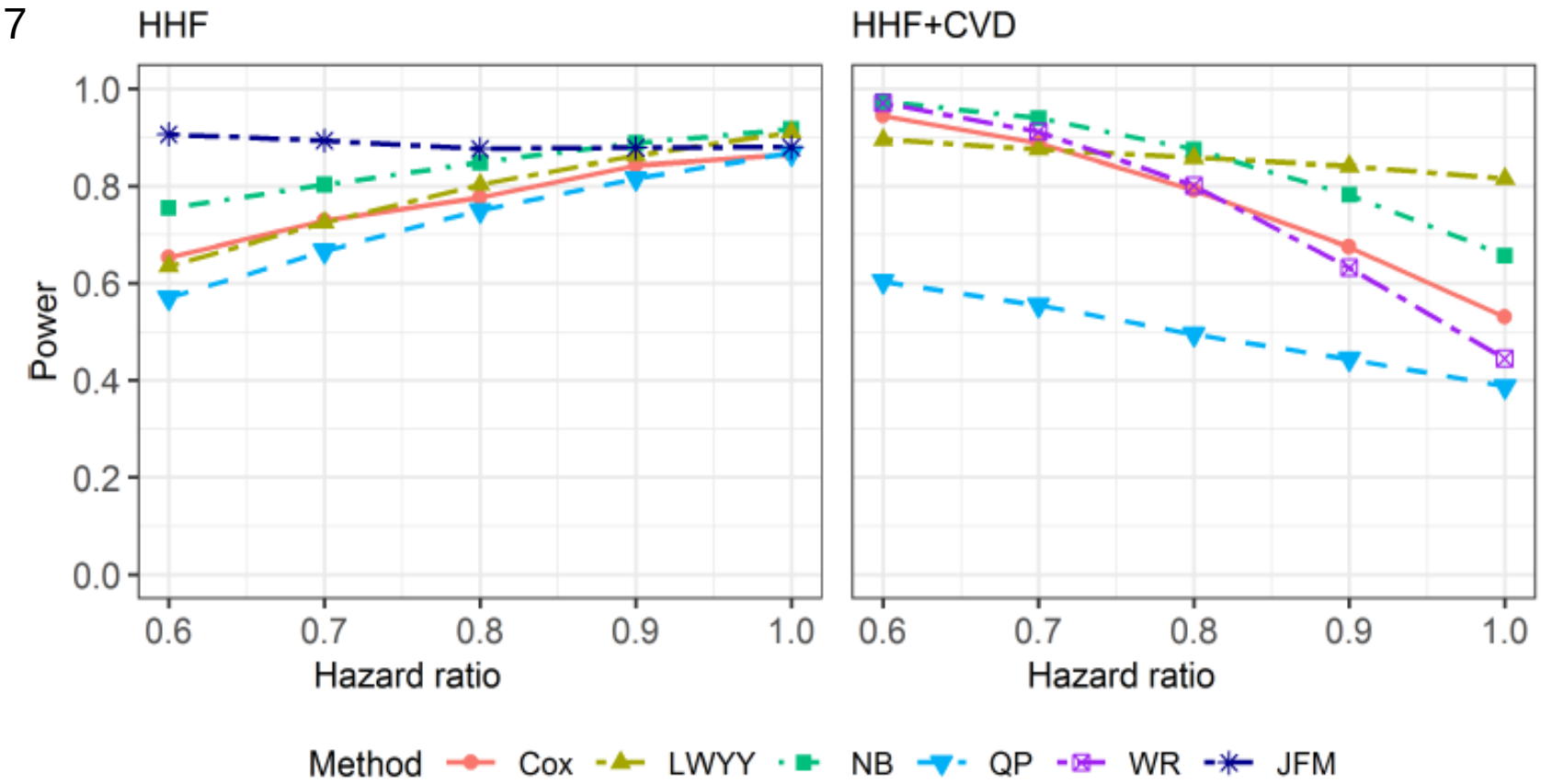
Simulation Study – Quasi-Poisson with Fletcher‘s Method (2)

Table 1: Type I error rates (1-sided tests, nominal significance level $\alpha = 0.025$) for HHF and HHF+CVD and RR=1 including quasi-Poisson based on Fletcher's estimate of ϕ

Endpoint	HR	LWYY	NB	Cox	JFM	Win ratio	Quasi-Poisson	Quasi-Poisson (Fletcher)
HHF	0.6	0.000	0.004	0.005	0.024	-	0.000	0.001
	0.8	0.005	0.010	0.010	0.019	-	0.003	0.007
	1.0	0.023	0.025	0.023	0.028	-	0.015	0.033
	1.25	0.100	0.057	0.058	0.031	-	0.067	0.127
HHF+CVD	1.0	0.024	0.024	0.022	-	0.024	0.002	0.072
	1.25	0.023	0.004	0.004	-	0.002	0.001	0.069

Simulation Study – 20% Drug Disc. after each HHF

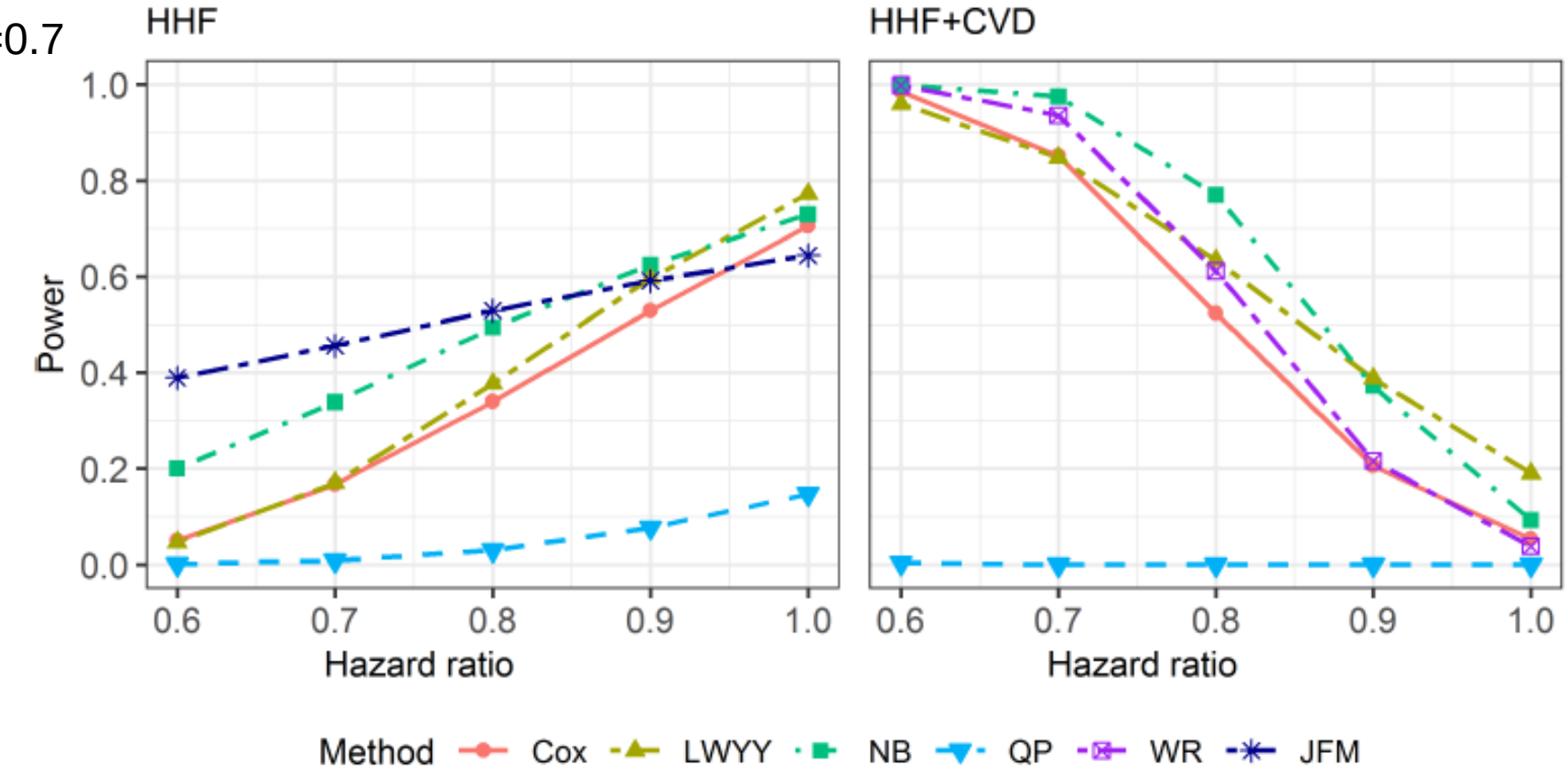
RR=0.7



- Clagett et al. (2018) also showed that recurrent event methods have higher power, if drug disc. after first event ≤ 30%

Simulation Study – Increased CV Mortality

RR=0.7



- CV mortality in study increased from 12.5% to to 40%
  - Behaviour from base case more pronounced for both endpoints