UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

EXPLORATORY AND CONFIRMATORY SUBGROUP ANALYSES IN CLINICAL TRIALS

EFSPI Biomarkers and Subgroups Meeting

Leiden, 24 June 2016

Tim Friede

Department of Medical Statistics

University Medical Center Göttingen

ACKNOWLEDGEMENTS

UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

- **“Improving outcomes from the treatment of low back pain”** funded by UK NIHR (RP-PG-0608-10076); PI Martin Underwood (Warwick)
- **“BI**ostatistische **M**ethoden zur effizienten Evaluation von **I**ndividualisierten **T**herapien **(BIMIT)”** funded by BMBF
  - Coordinated by Meinhard Kieser (Heidelberg)
  - WP C: Tim Friede**,** Marius Placzek, Roland Gera (Göttingen); Heinz Schmidli (Novartis)

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

BIMIT

- **“Identification und confirmation of biomarker-defined populations in the personalized pharmacotherapy”** co-funded by BfArM
  - PIs Tim Friede, Jürgen Brockmöller (UMG), Norbert Benda Julia Stingl (BfArM)
- **"Innovative methodology for small populations research" (InSPiRe)** funded by EU's FP7 (HEALTH 2013 – 602144)
  - Coordinated by Nigel Stallard (Warwick)
  - WP4: Tim Friede, Christian Röver, Steffen Unkel (Göttingen); Beat Neuenschwander, Simon Wandel (Novartis); …

InSPiRe
Innovative methodology for small populations research

OUTLINE

- **Motivation**: Biomarkers, Personalised medicine
- **Identifying subgroups** in a single trial
- Extension to several trials: **Meta-analytic framework**
- **Clinical development plans**: Integration of subgroup identification and confirmation
- **Concluding remarks**

WHAT ARE BIOMARKERS?

- Definition by the **Biomarkers Definitions Working Group** (2001)
  - „A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.“
  - Very general definition
- **Sometimes more restrictive use of the term „biomarker“** relating only to measurements in blood, urine, CSF, …
- Excluding e.g. imaging markers

WHAT ARE BIOMARKERS USED FOR?

- Biomarkers are used …
  - to **diagnose** diseases (or certain subtypes)
  - to **predict** disease course or response to treatment
  - to **stratify** populations
  - to **monitor** patients
  - as **endpoints** in clinical trials

WHAT ARE BIOMARKERS USED FOR?

- Biomarkers are used …
  - to **diagnose** diseases (or certain subtypes)
  - to **predict** disease course or response to treatment
  - to **stratify** populations
  - to **monitor** patients
  - as **endpoints** in clinical trials

PROGNOSTIC VS. PREDICTIVE BIOMARKERS

- **prognostic** biomarkers
  - markers affecting disease course

- **predictive** biomarkers
  - markers affecting treatment effect
  - technically: interactions with treatment effect, also known as moderators

- **Not necessarily the same biomarkers!**

PREDICTIVE BIOMARKERS

- **Personalised medicine**
  - Efficacy, safety and consequently benefit-risk might vary across patient population
  - Stratification of patient populations
  - Drive towards targeted treatments

- **Enrichment** of clinical study populations (Temple, 2010)
  - “to identify a population of patients in whom a drug effect, if present, is more likely to be demonstrable”
  - (a) practical, (b) prognostic, and (c) predictive enrichment

STRATIFIED MEDICINE: EXAMPLES OF TARGETED THERAPIES

Table I. Oncology products approved in the USA for selected populations.

Compound	Target	Indication
Crizotinib (Xalkori®)	ALK	ALK-rearranged non-small cell lung cancer
Vemurafenib (Zelboraf®)	BRAF	BRAF mutant advanced melanoma
Trametinib (Mekinist®)	MEK	BRAF mutant advanced melanoma
Trastuzumab (Herceptin®)	Her 2	Her 2 expressing breast cancer
Lapatinib (Tykerb®)	Her 2	Her 2 expressing metastatic gastric cancer
Rituximab (Rituxan®)	CD20	CD20(+) B-cell lymphomas
Cetuximab (Erbitux®)	EGFR	$KRAS^{wt}$, EGFR(+) metastatic colorectal cancer
Panitumumab (Vectibix®)	EGFR	$KRAS^{wt}$, EGFR(+) metastatic colorectal cancer

Table I from Mehta et al. (2014) Stat Med

EXAMPLE: IMPROVING OUTCOMES FROM THE TREATMENT OF LOW BACK PAIN

- NIHR funded project lead by Martin Underwood (Warwick, UK)
- **Project aim**
  - „… to improve the clinical and cost-effectiveness of low back pain treatment by providing patients, their clinical advisors, and health service purchasers with better **information about which participants are most likely to benefit from which treatment choices.**"
- **Repository**
  - Individual patient data of 19 randomised controlled trials
  - Total of 9,328 patients

OUTLINE

- **Motivation**: Biomarkers, Personalised medicine
- **Identifying subgroups** in a single trial
- Extension to several trials: **Meta-analytic framework**
- **Clinical development plans**: Integration of subgroup identification and confirmation
- Concluding remarks

MODERATORS OF TREATMENT EFFECT

- Baseline variables affecting treatment effect; sometimes also referred to as **“predictive” factors** (not to be confused with prognostic factors)
- Technically **interaction effects** between baseline variable and treatment effect
- For instance, analysis of covariance (ANCOVA) with treatment, baseline covariables and treatment-by-baseline covariable interactions
- **More sophisticated**: Fractional polynomials (Royston & Sauerbrei, 2004)

SUBGROUP IDENTIFICATION

- For an overview refer to recent systematic literature **review by Ondra et al. (2015) on methods for subgroup identification and confirmation** in clinical trials
- **Exploratory subgroup identification**
  - attracted a lot of attention over the past years
  - several methods proposed
- Here we describe one we adopted when working on the back pain repository …

ADAPTIVE REFINEMENT BY DIRECTED PEELING (ARDP) ALGORITHM

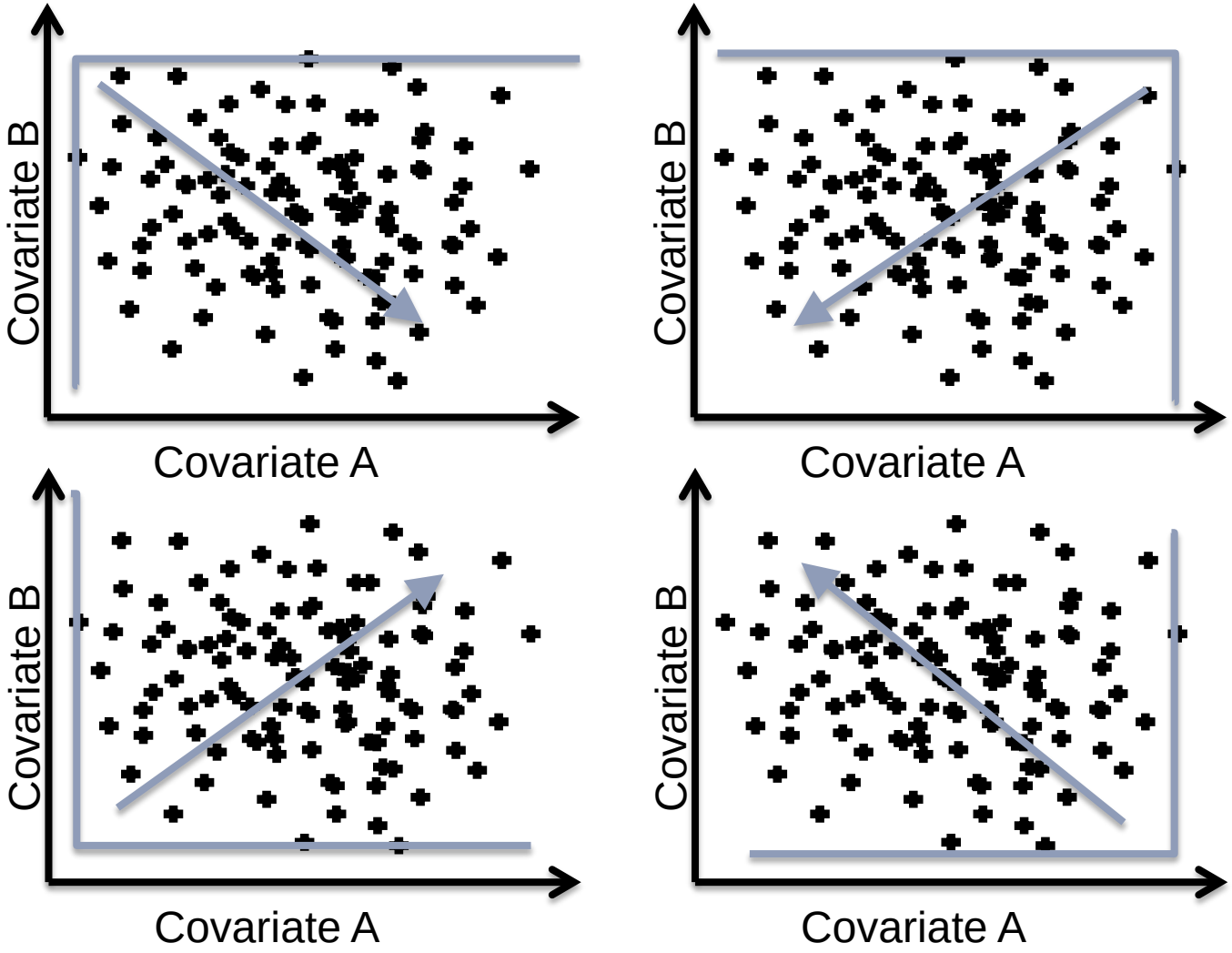
- Proposed by LeBlanc et al. (2005) to identify risk groups (**prognostic factors**)
- Risk groups defined by (half open) **“boxes”** resulting in simple rules
- Here modified to identify subgroups responding particularly well to treatment (**predictive factors**)

SUB-GROUP IDENTIFICATION: ADAPTIVE REFINEMENT BY DIRECTED PEELING

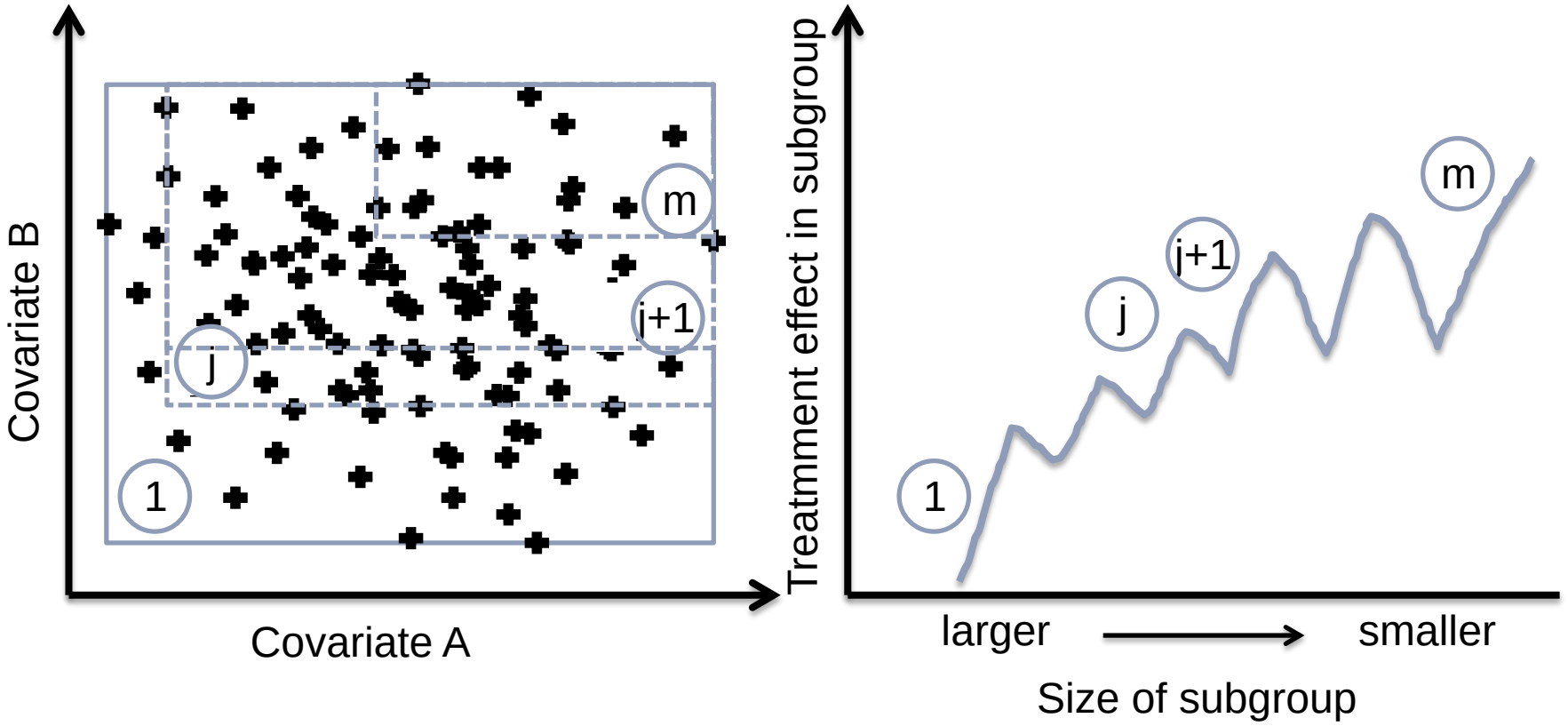
UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

1. Investigating **interactions of covariates with treatment** determines covariates to be included and direction of peeling
2. Start with a “subgroup” B^0 that includes all observations.
3. For each variable we **peel a certain number of observations off** resulting in subgroups $B_j^m, j = 1, \dots, p$.
4. For each subgroup B_j^m calculate the **treatment-by-subgroup interaction** and select the B_j^m which gives the largest improvement on the interaction effect in comparison to the previous iteration. The selected subgroup is then called B^{m+1}.
5. Estimate the treatment effects for the outcome of interest for subgroup B^{m+1}.
6. Repeat steps 3 to 5 until the size of the remaining region is not smaller than *r*.

SUB-GROUP IDENTIFICATION: ADAPTIVE REFINEMENT BY DIRECTED PEELING (ARDP)



SUB-GROUP IDENTIFICATION: ADAPTIVE REFINEMENT BY DIRECTED PEELING (ARDP)



SUB-GROUP IDENTIFICATION: ADAPTIVE REFINEMENT BY DIRECTED PEELING (ARDP)

- **Algorithm can be applied to various kinds of endpoints**
  - Continuous: Gaussian linear models
  - Binary: logistic regression
  - Time-to-event: Cox proportional hazard models
- **No distributional assumption regarding the covariates** required, but they should be ordinal with sufficient number of possible outcomes
- If covariable not ordinal, then order could be imposed: order the categories by the regression coefficients estimated in Step 1 of the algorithm (LeBlanc et al., 2005).

SUB-GROUP IDENTIFICATION: ADAPTIVE REFINEMENT BY DIRECTED PEELING (ARDP)

- „Experience with simulated data with low signal shows that there can be **substantial estimation bias due to peeling** if there are a moderate number of predictors (p>5).“ (LeBlanc et al., 2005)
- LeBlanc et al. (2005) suggested **resampling methods** to reduce selection bias and for inference
- **K-fold crossvalidation** to reduce bias in estimation
- **Permutation test** to test whether the prognostic subgroups are associated with outcome

OUTLINE

- **Motivation**: Biomarkers, Personalised medicine
- **Identifying subgroups** in a single trial
- Extension to several trials: **Meta-analytic framework**
- **Clinical development plans**: Integration of subgroup identification and confirmation
- Concluding remarks

MODERATORS OF TREATMENT EFFECT

- Modelling **between-study heterogeneity**
- **Hierarchical (mixed-effects) model**
  - Fixed effects: treatment, covariables, treatment-by-covariable interactions
  - Random effects: trial and trial-by-treatment interaction (a model similar to standard random-effects meta-analysis)
  - **Example with continuous outcome in SAS**

```
1 proc mixed data=&data;
2    class &trt &trials;
3    model &outcome = &trt &var &trt*&var / s ddfm=satterth;
4    random intercept &trt / subject=&trials;
5    repeated / group=&trials;
6 run;
```

SUB-GROUP IDENTIFICATION: ADAPTIVE REFINEMENT BY DIRECTED PEELING (ARDP)

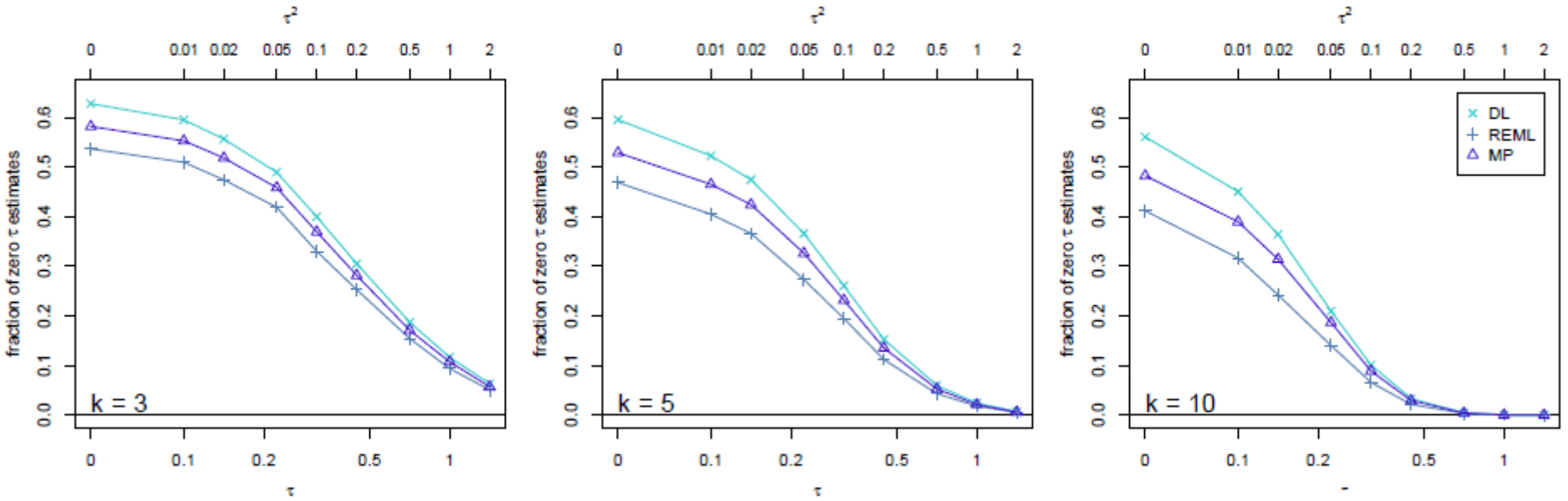
- Extension to multiple trials by including terms for between-trial heterogeneity in the model
- **Random effects meta-analyses of interaction effects**
  - Two-step procedure: interaction effects estimated from individual trials are combined in random-effects meta-analyses
  - One-step procedure: hierarchical model

BETWEEN-TRIAL HETEROGENEITY

- Likely to be present due to some differences in e.g. trial populations (see e.g. Higgins et al, 2009)
- Variety of estimators proposed including REML, MoM / DL, PM, … (see Veronicki et al, 2015)
- Estimation particularly challenging with only few studies (a situation frequently encountered)
- In the following: some results for pairwise meta-analysis with few small studies motivated by rare disease setting (InSPiRe)

ESTIMATION OF BETWEEN-TRIAL HETEROGENEITY

- Proportion of between-trial heterogeneity estimates being 0
- Estimators: DerSimonian-Laird (DL), restricted maximum likelihood (REML), Mandel-Paule (MP)

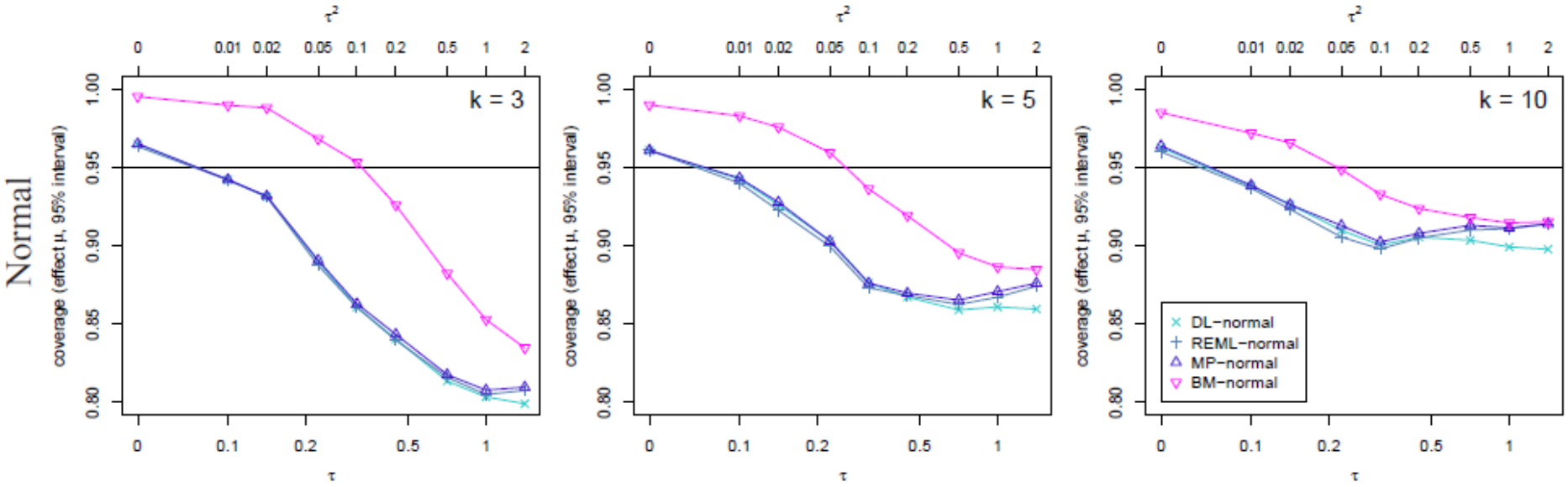


Friede et al. (2016) RSM

InSPiRe
Innovative methodology for small populations research

ESTIMATION OF BETWEEN-TRIAL HETEROGENEITY

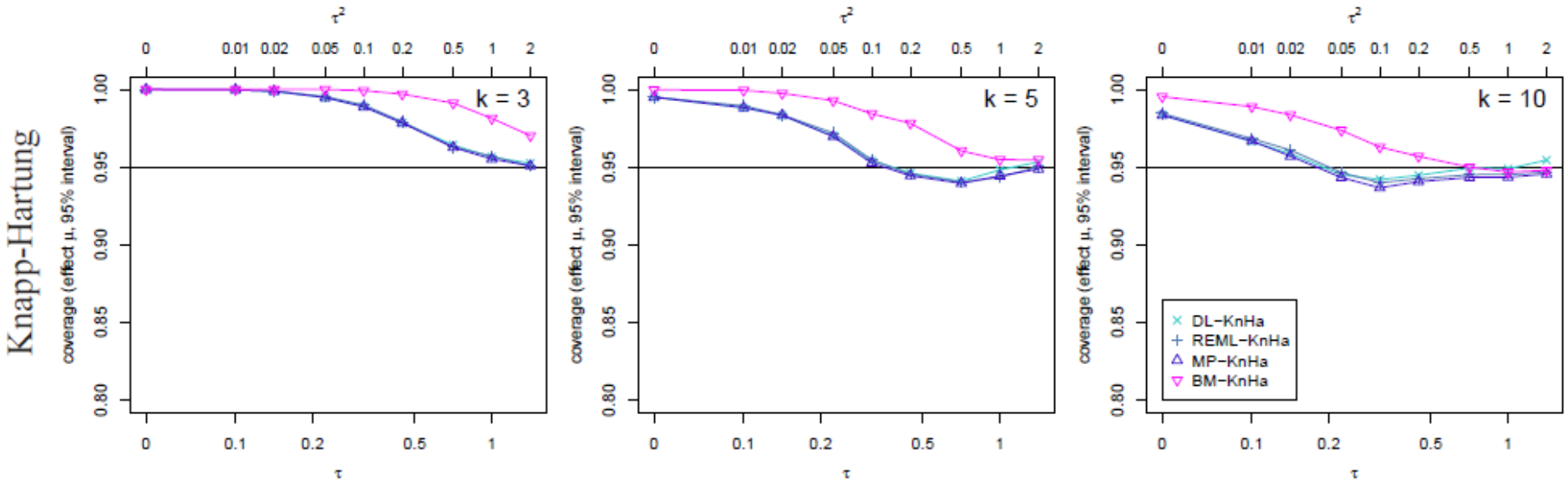
- Coverage probability for confidence intervals of combined effect
- Construction of confidence intervals using normal quantiles
- Estimators: DerSimonian-Laird (DL), restricted maximum likelihood (REML), Mandel-Paule (MP), **Bayes-modal** (BM)



Friede et al. (2016) RSM

ESTIMATION OF BETWEEN-TRIAL HETEROGENEITY

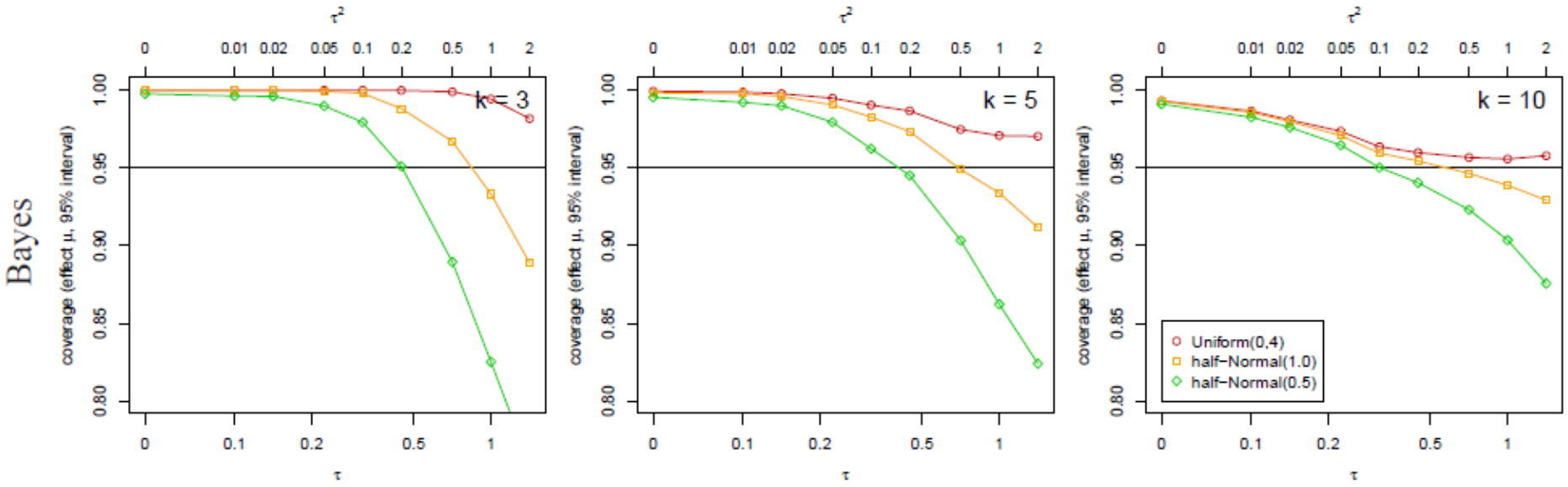
- Coverage probability for confidence intervals of combined effect
- Construction of confidence intervals using **Knapp-Hartung method** (using t-quantiles and scaling of standard error)



Friede et al. (2016) RSM

ESTIMATION OF BETWEEN-TRIAL HETEROGENEITY

- Coverage probability for credibility intervals of combined effect
- Bayes with “weakly informative” priors for tau
- R package **`bayesmeta`** available on CRAN (Christian Röver)

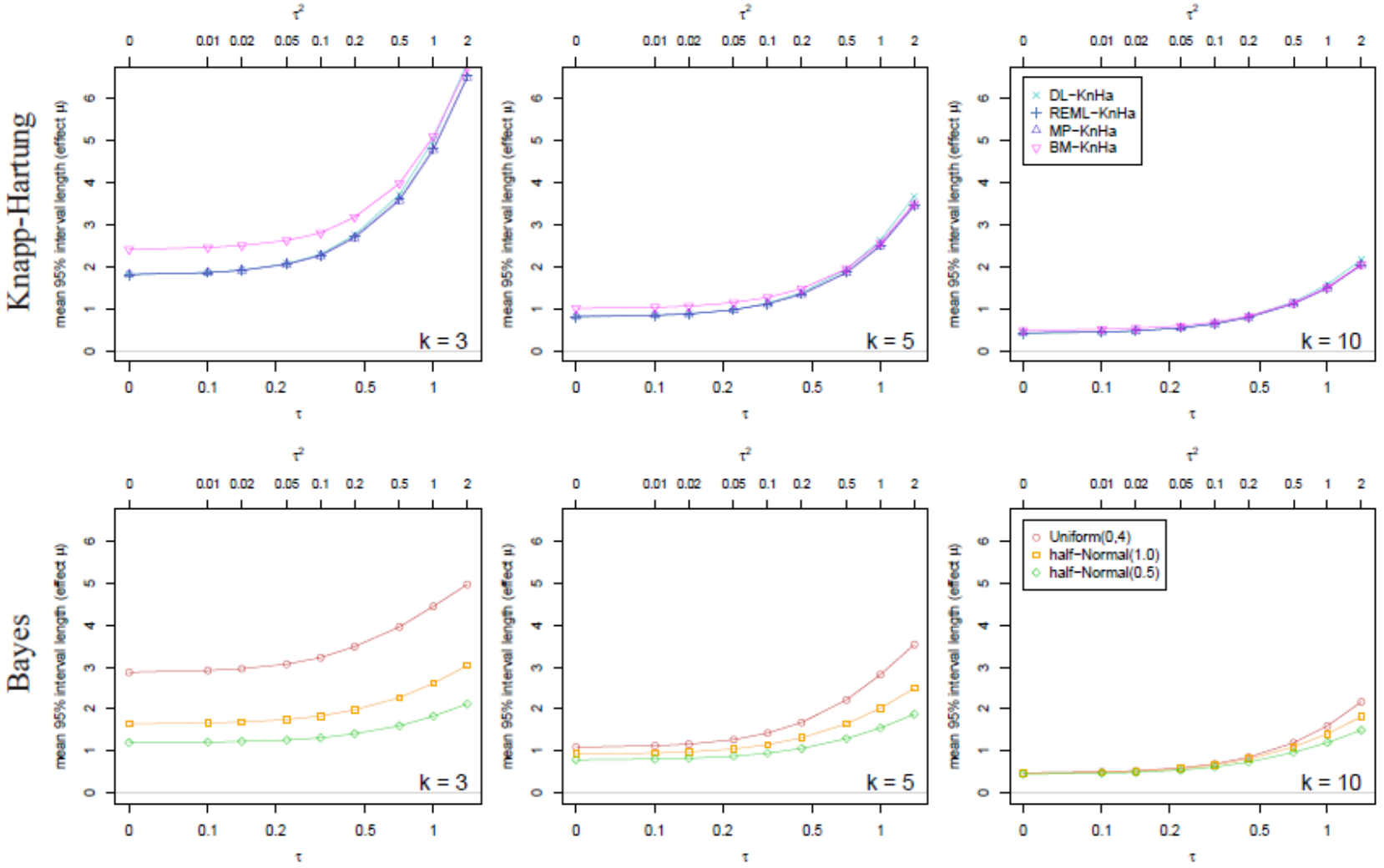


Friede et al. (2016) RSM

BETWEEN-TRIAL HETEROGENEITY

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

- Mean length of confidence / credibility intervals

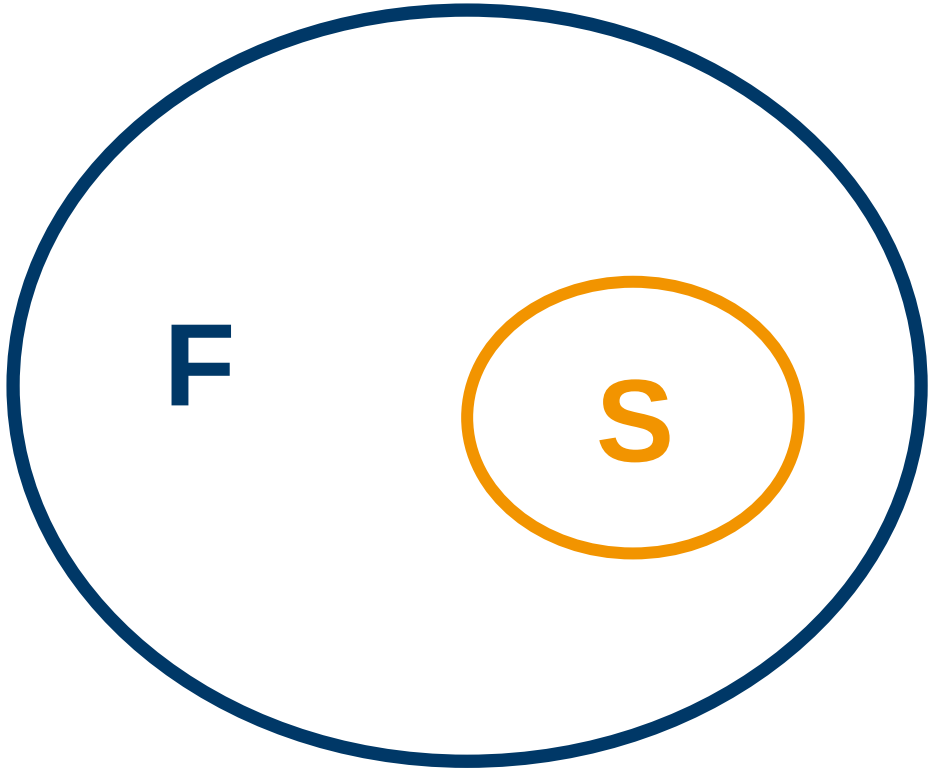


Friede et al. (2016) RSM

OUTLINE

- **Motivation**: Personalised medicine
- **Identifying subgroups** in a single trial
- Extension to several trials: **Meta-analytic framework**
- **Clinical development plans**: Integration of subgroup identification and confirmation
- Concluding remarks

STRATIFIED MEDICINE

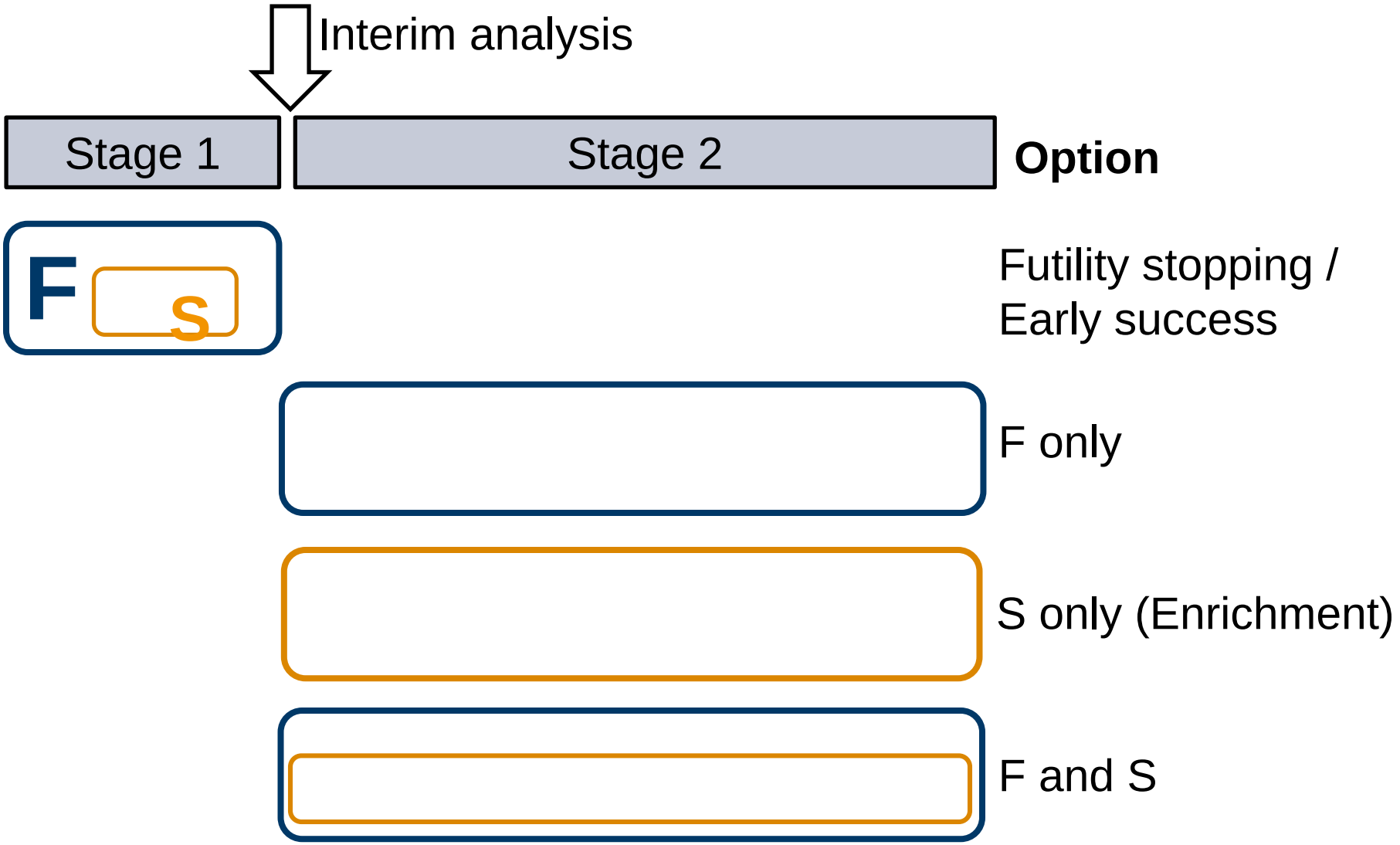


Situation considered

- Biomarker-defined subgroup identified in exploratory study
- Subgroup to be confirmed by independent data
- Confirmation of treatment effect in selected population

Full population **S**ub-population

ADAPTIVE ENRICHMENT DESIGN



STATISTICAL METHODOLOGY

- **repeated testing**
  - classical group sequential designs (e.g. Jennison & Turnbull 1999)
- **combining pre/ post adaptation data**
  - (recursive) combination test (Brannath et al, 2002), conditional error function approach (Müller & Schäfer, 2001)
- **multiple hypotheses**
  - closed test principle (Marcus et al, 1976), Bonferroni, . . .
- **combinations of these approaches in ASDs:** e.g. weighted inverse normal method and closed test principle

$$C(p_{S,1}, p_{S,2}) = 1 - \Phi(w_1 \Phi^{-1}(1 - p_{S,1}) + w_2 \Phi^{-1}(1 - p_{S,2}))$$

ENRICHMENT DESIGNS MORE POWERFUL

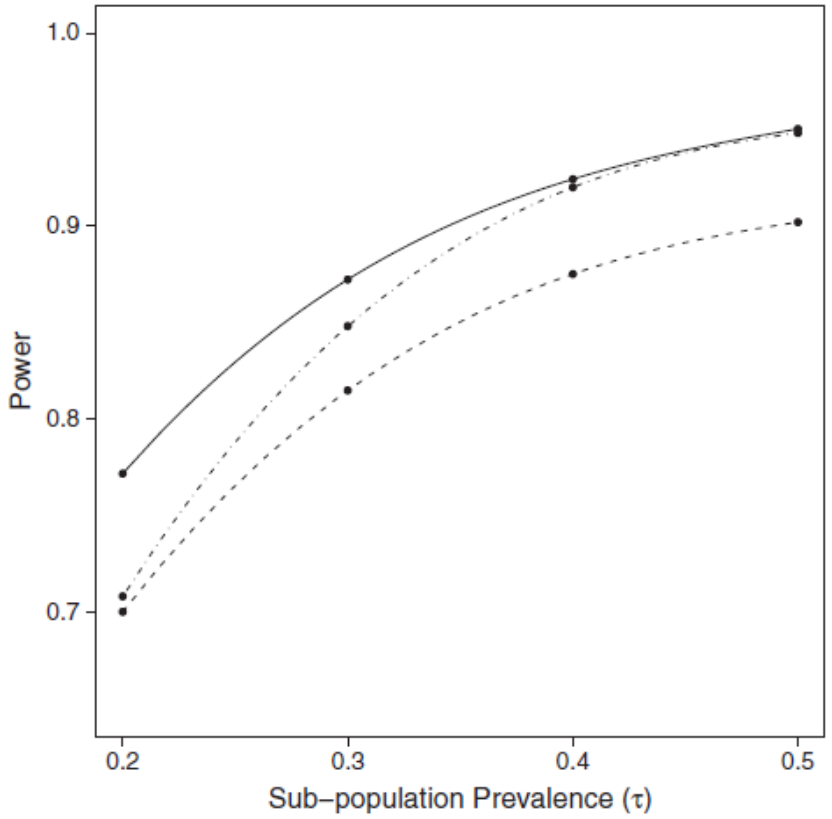


Figure 3. Power of rejecting at least one hypothesis of the adaptive conditional error function approach (——), separate designs (– – –) and the fixed design (– · –) for hazard ratios 0.77 in the subpopulation and 1 in the complement to the subpopulation (10 000 simulation replications per scenario).

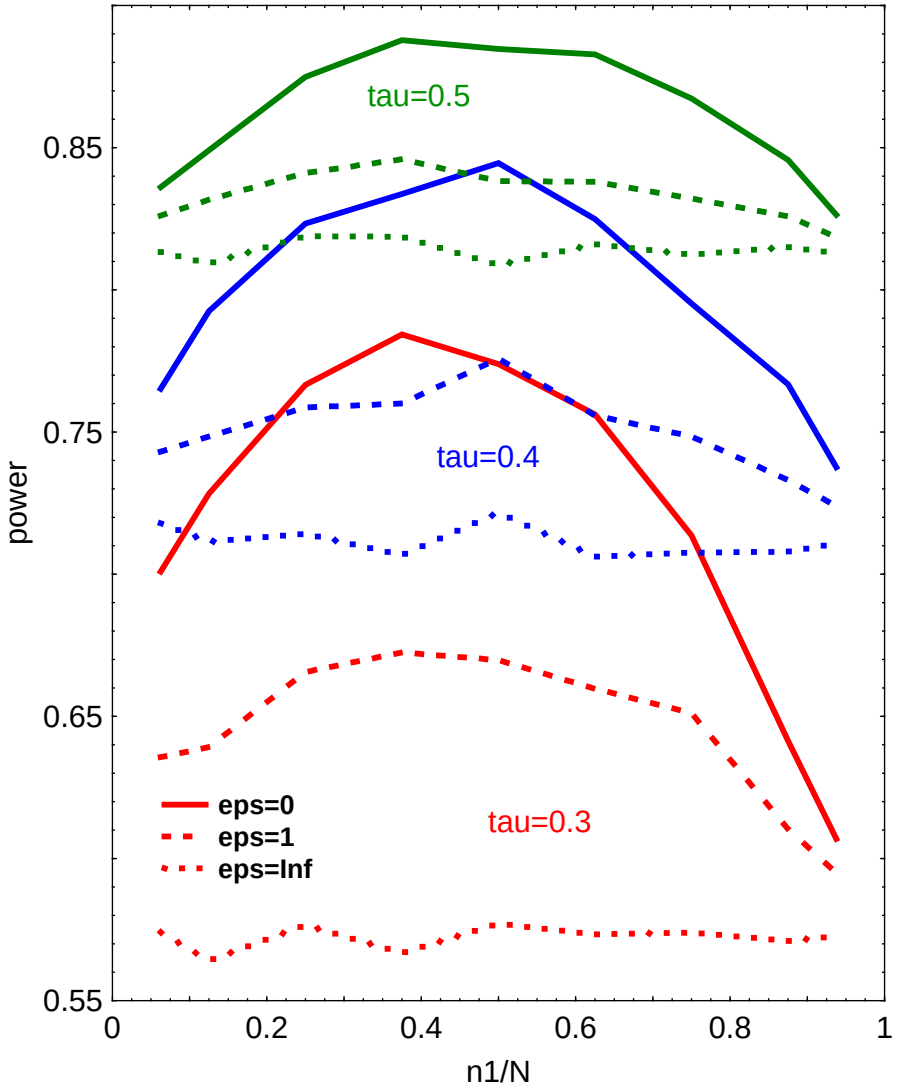
Friede et al. (2012) Stat Med

OPTIMAL TIME POINT FOR INTERIM ANALYSIS

- Adaptive enrichment design with CEF approach
- Simulation results for n_{sim}=10,000 replications
- n=400 subjects per group (treatment/placebo)
- Under the alternative

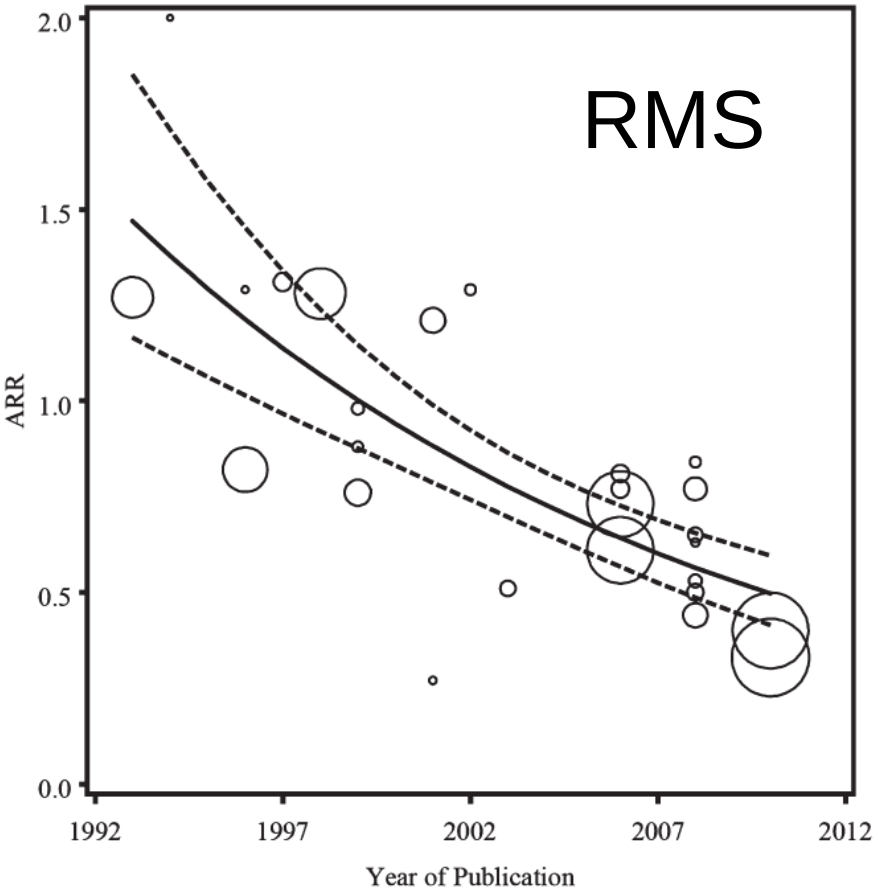
$$\Delta_{F \setminus S} = 0, \ \Delta_S = 0.3$$

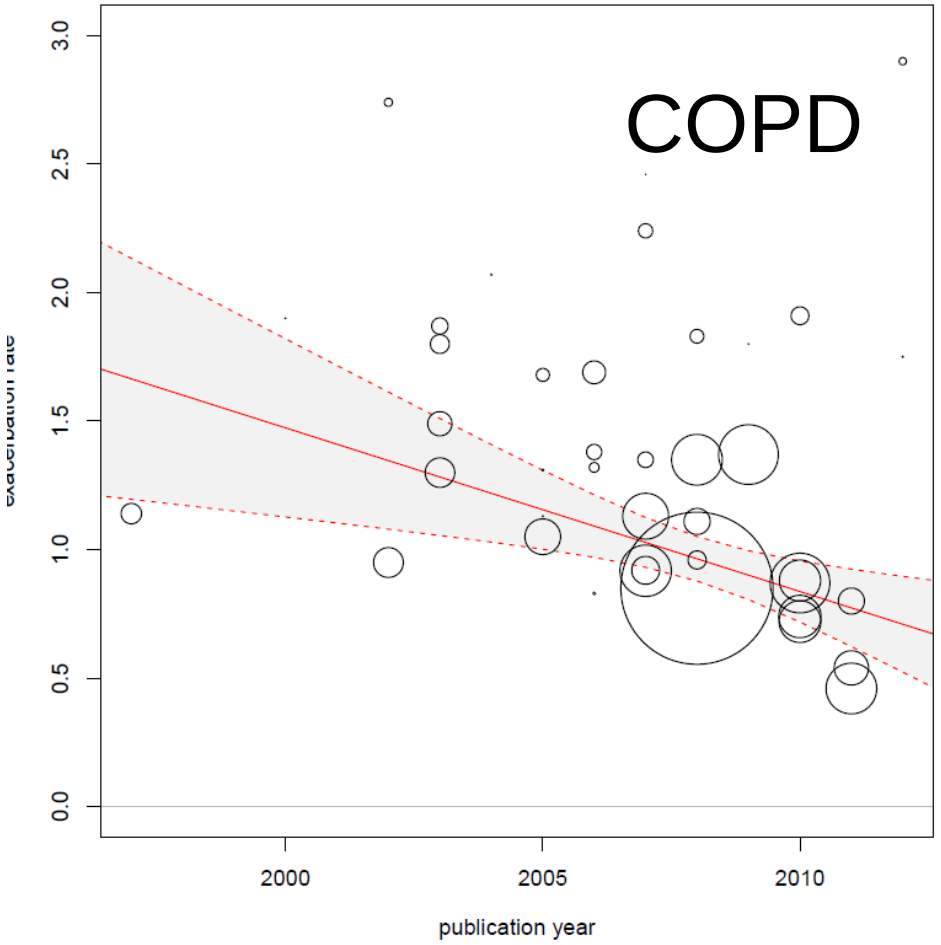
- Maximum in power after 40-50% of the subjects



UNCERTAINTY IN PLANNING TRIALS: Trends in Placebo Event rates in Chronic Conditions

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

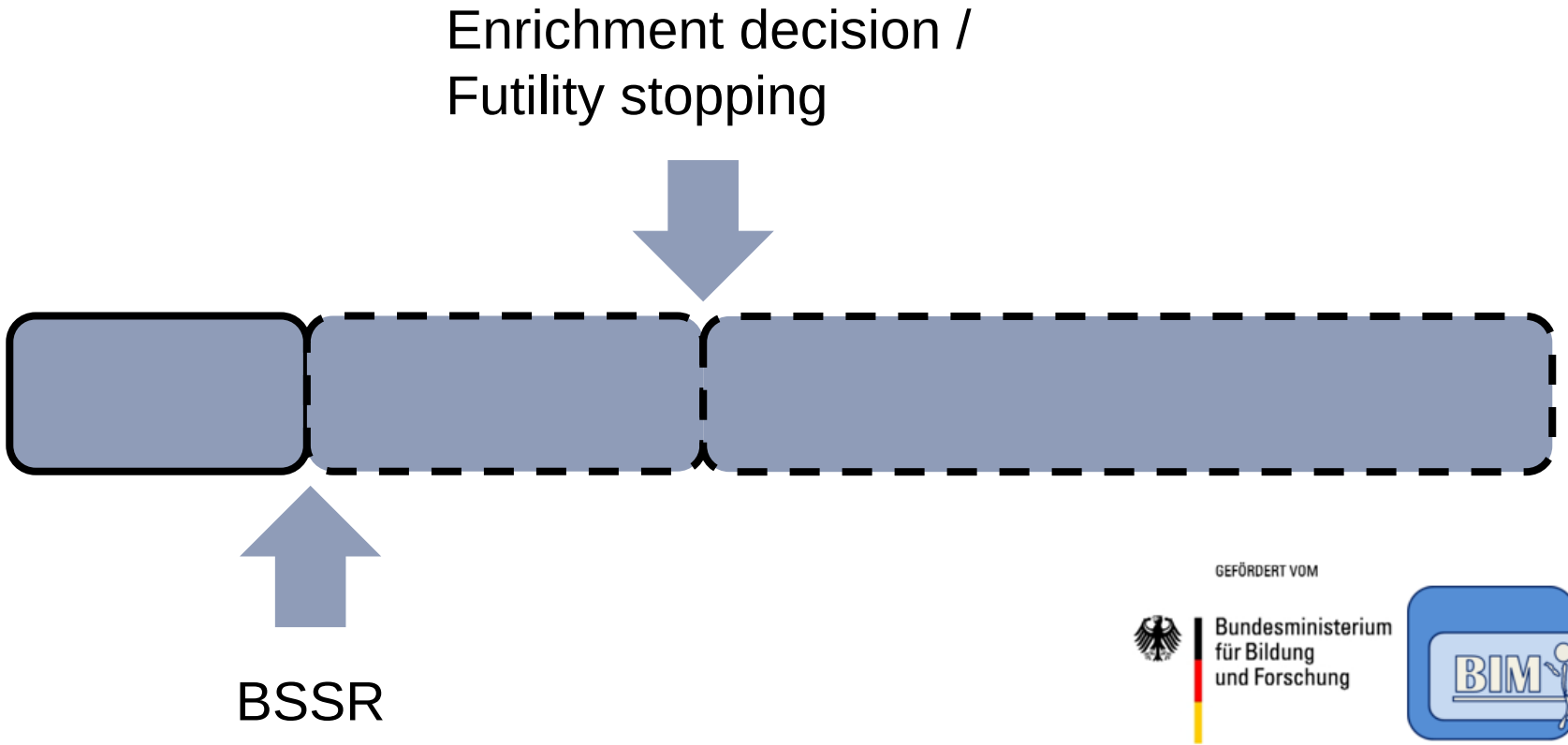




Nicholas et al. (2011) MSJ

Andreas, Röver et al. (2016)

BLINDED SAMPLE SIZE REESTIMATION (BSSR) IN ADAPTIVE ENRICHMENT DESIGNS

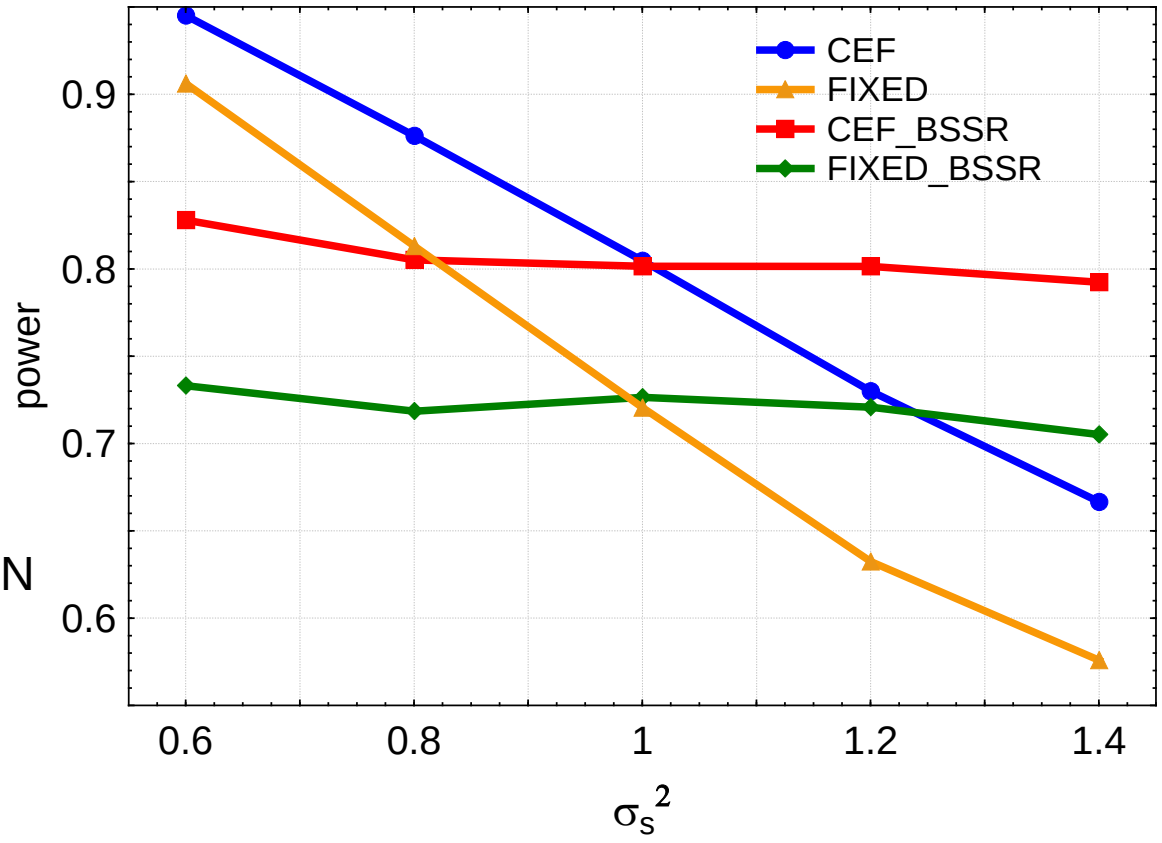


GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
BIMT

- Early IA for **blinded** sample size reestimation
- Later IA for enrichment decision / futility stopping (**unblinding**)

BSSR IN ADAPTIVE ENRICHMENT DESIGNS: SIMULATION STUDY

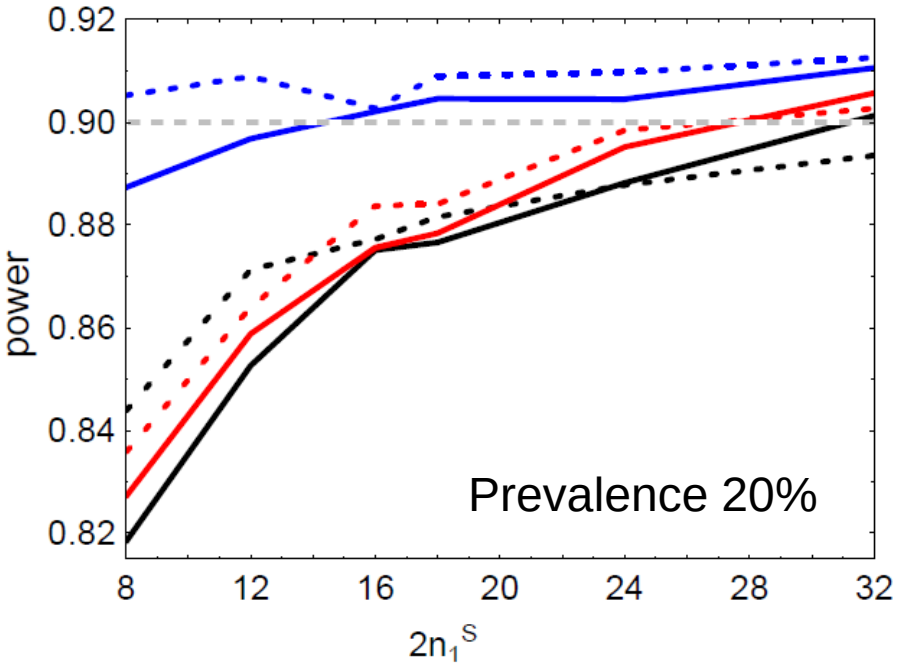
- n_{sim}=10,000, $\tau = 0.5$
- $\alpha = 0.025,\ 1-\beta = 0.8$
- $\Delta_{F\setminus S} = 0,\ \Delta_S = 0.5$
- $\sigma_F^{2*} = \sigma_F^2 = 1,\ \sigma_S^{2*} = 1$
- BSSR at 30% of N_0
- Interim Analysis at 50% of N ($\varepsilon = 1$)

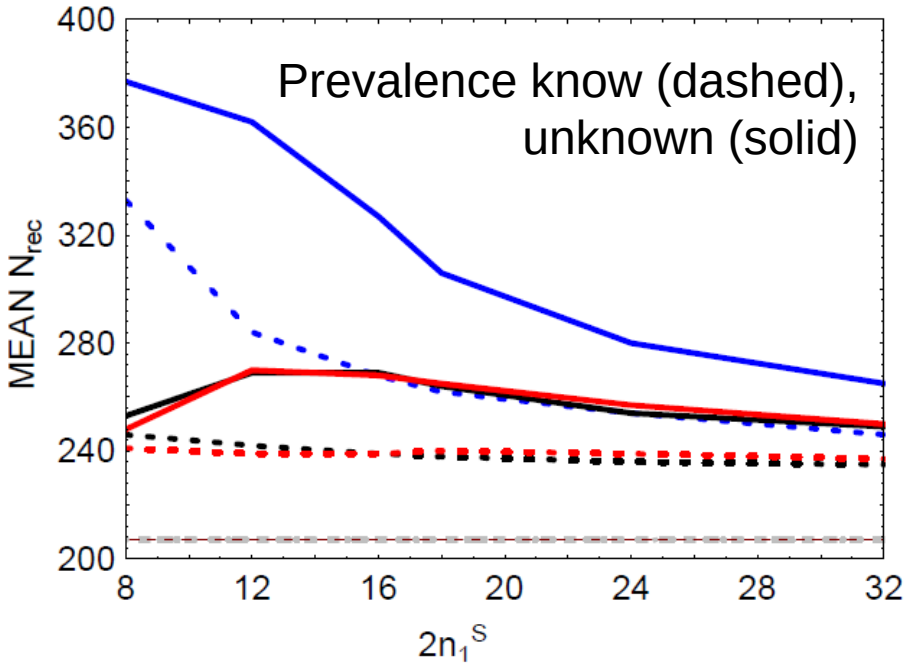


BSSR IN STUDY WITH SUBGROUP

- Simulated power and corresponding recalculated sample sizes depending on the number of subjects in the subgroup at the timepoint of the blinded review

$$1-\beta = 0.9 \quad \Delta_S = 1 \quad \sigma_S = 1.3 \quad n_{sim} = 10,000$$





- With small sample sizes / subgroups: uncertainty in estimation of nuisance parameters has to be accounted for leading to large average sample sizes

Placzek & Friede (2016)

OUTLINE

- **Motivation**: Biomarkers, Personalised medicine
- **Identifying subgroups** in a single trial
- Extension to several trials: **Meta-analytic framework**
- **Clinical development plans**: Integration of subgroup identification and confirmation
- Concluding remarks

Clinical Scenario Evaluation (CSE)

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Framework for the Assessment of Competing Strategies

(Benda et al (2010) DIJ; Friede et al (2010) DIJ)

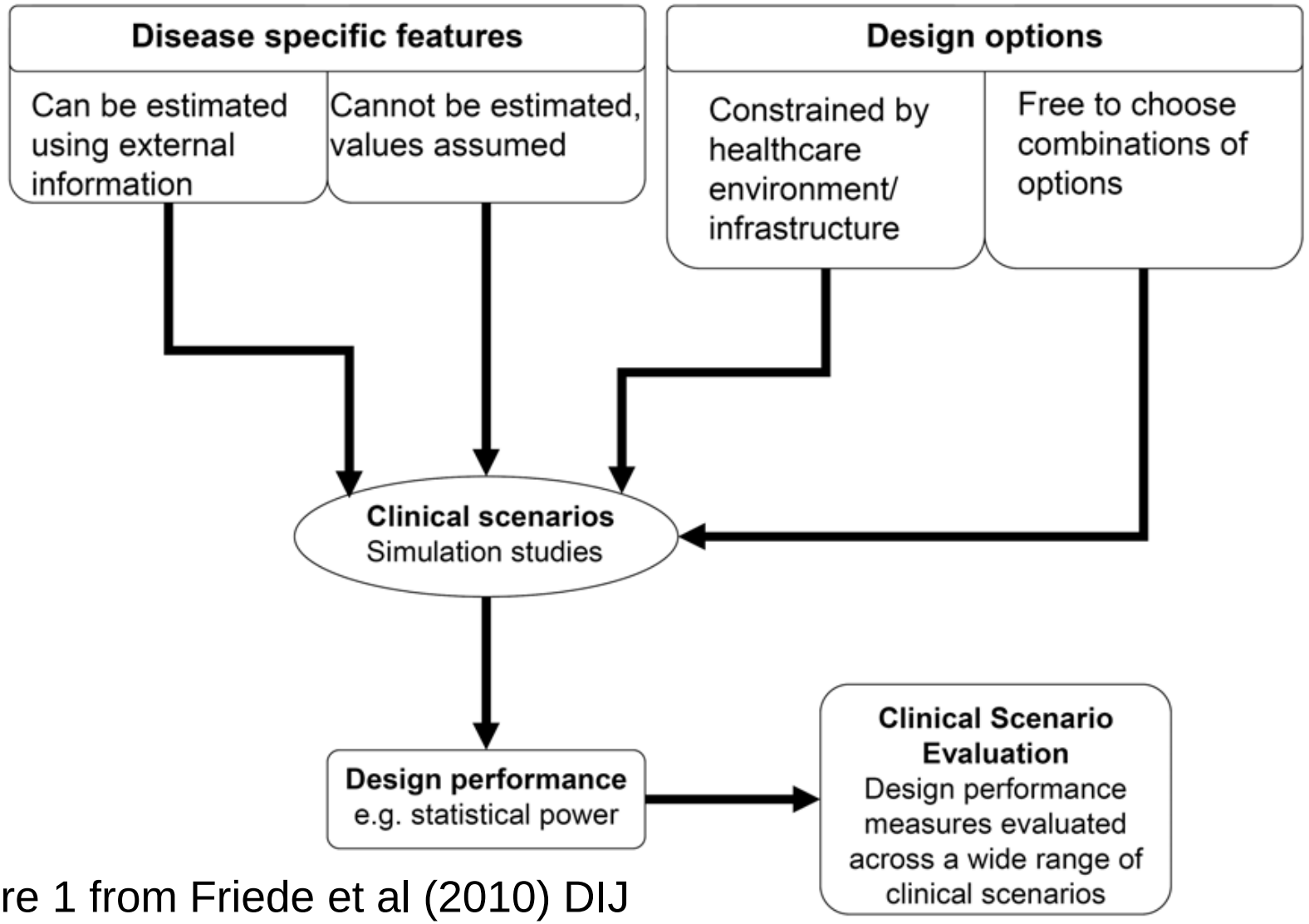


Figure 1 from Friede et al (2010) DIJ

CONCLUDING REMARKS

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

- **Subgroup identification based on Adaptive Refinement by Directed Peeling (ARDP)**
  - Facilitates decision making on subgroup selection balancing size of subgroup with size of treatment effect
- **Subgroup identification from multiple trials**
  - Some level of between-trial heterogeneity expected and should be reflected in statistical model
  - Estimation difficult if only a small number of studies included in the analysis
- **Gain in power by adaptive enrichment design** compared to separate studies / fixed design can be substantial
- **Assessment of complex development plans usually requires extensive simulations**