UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

CLINICAL REGISTRIES
Use and Emerging Best Practices

Tim Friede

Department of Medical Statistics

University Medical Center Göttingen

DZHK (German Center for Cardiovascular Research)

DZHK
DEUTSCHES ZENTRUM FÜR
HERZ-KREISLAUF-FORSCHUNG E.V.

OUTLINE

- Background: Definition(s) and classification
- Purposes of clinical registries
- Statistical issues and methodologies
- Operational and logistical issues
- Conclusions and discussion

DEFINITION(S) OF CLINICAL REGISTRIES

- **Not one, but many definitions in use**
- Also called **patient registries, clinical data registries, disease registries, outcomes registries**, …
- "… a file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way, in order to serve a predetermined purpose." (Brooke, 1974)
- "… an organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons who have either a particular disease, a condition (e.g., a risk factor) that predisposes [them] to the occurrence of a health related event, or prior exposure to substances (or circumstances) known or suspected to cause adverse health effects." (US National Committee on Vital and Health Statistics)

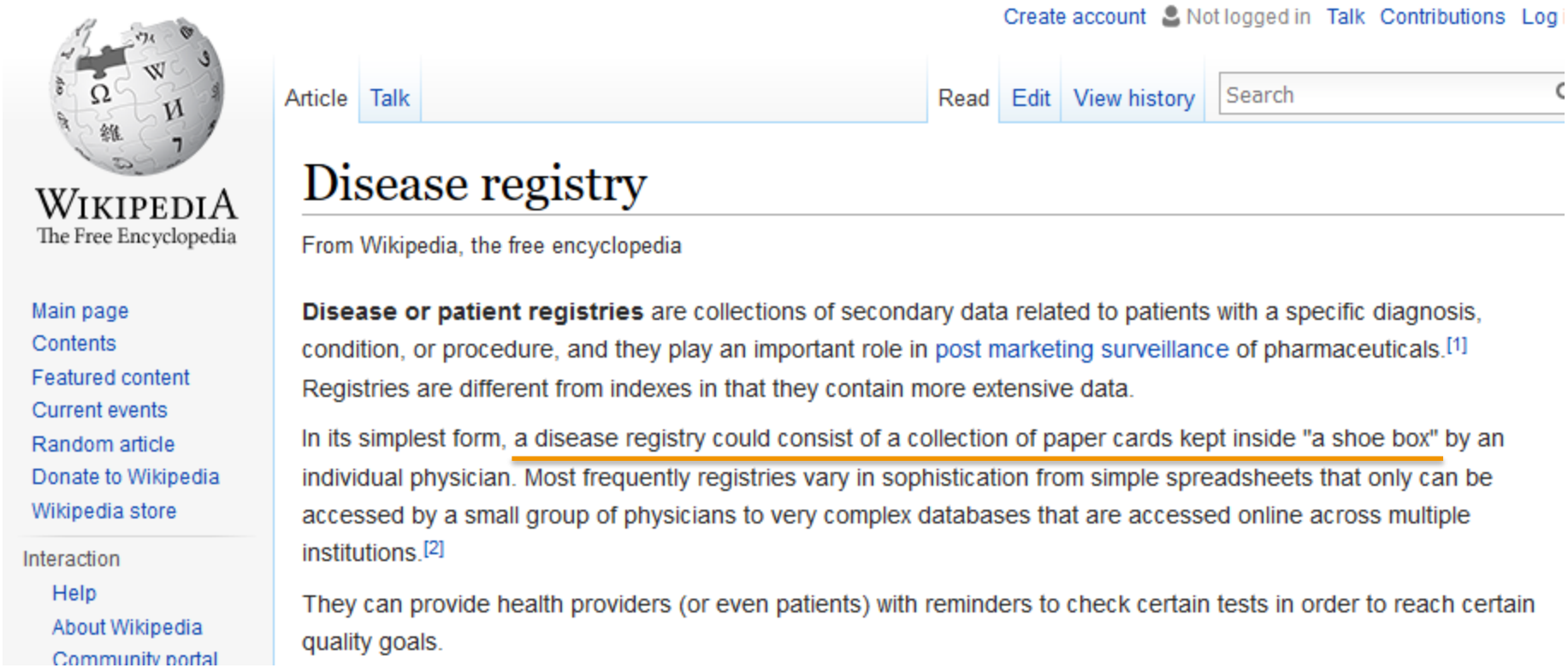
Classification by the way the population is defined

DEFINITION(S) OF CLINICAL REGISTRIES

- "… an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes. A registry database is a file (or files) derived from the registry."

- **Reference**: Gliklich & Dreyer eds. (2010) Registries for Evaluating Patient Outcomes: A User's Guide. (Available online!)

DEFINITION(S) OF CLINICAL REGISTRIES



Create account Not logged in Talk Contributions Log

WIKIPEDIA
The Free Encyclopedia

Article Talk Read Edit View history Search

Disease registry

From Wikipedia, the free encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal

Disease or patient registries are collections of secondary data related to patients with a specific diagnosis, condition, or procedure, and they play an important role in post marketing surveillance of pharmaceuticals.[1] Registries are different from indexes in that they contain more extensive data.

In its simplest form, a disease registry could consist of a collection of paper cards kept inside "a shoe box" by an individual physician. Most frequently registries vary in sophistication from simple spreadsheets that only can be accessed by a small group of physicians to very complex databases that are accessed online across multiple institutions.[2]

They can provide health providers (or even patients) with reminders to check certain tests in order to reach certain quality goals.

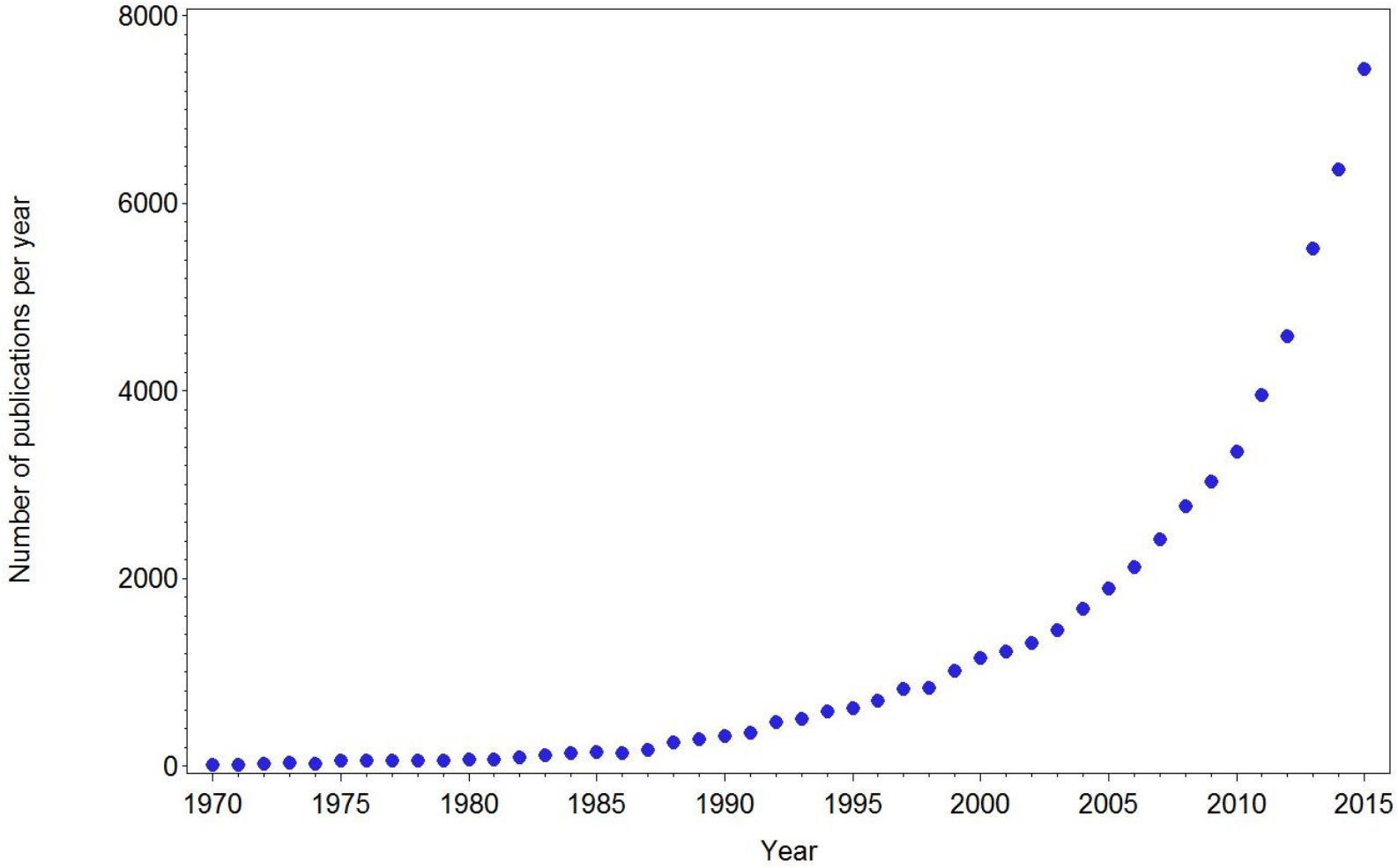
CLASSIFICATION BY THE WAY THE POPULATION IS DEFINED

- **Product registries**
  - drugs or medical devices
- **Health services registries**
  - patients who have had a common procedure, clinical encounter, or hospitalization
- **Disease or condition registries**
  - defined by patients having the same diagnosis

- **Reference**: Gliklich & Dreyer (2010)

PUBMED SEARCH

- **Search terms: disease registry OR clinical registry**



OUTLINE

- Background: Definition(s) and classification
- Purposes of clinical registries
- Statistical issues and methodologies
- Operational and logistical issues
- Conclusions and discussion

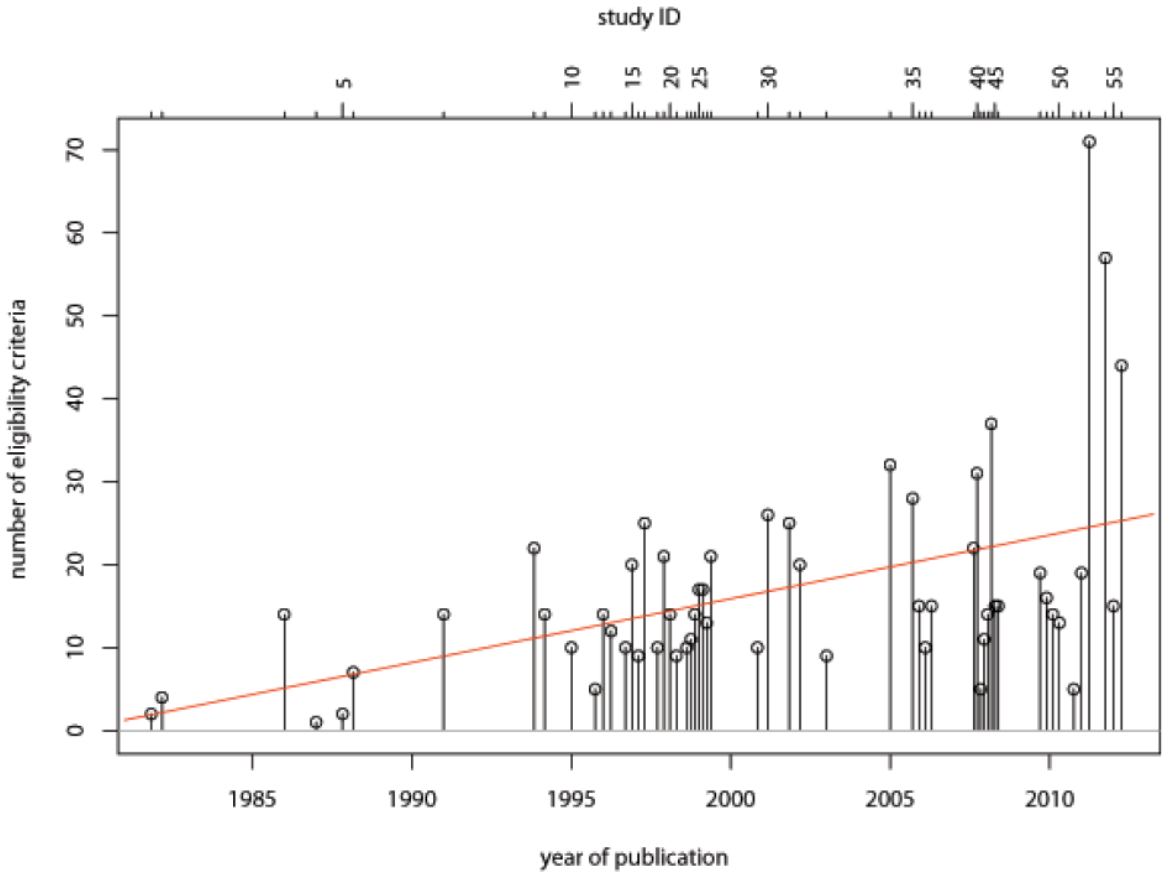
PURPOSES OF CLINICAL REGISTRIES

Purposes of clinical registries are manifold including …

- **Epidemiology**
  - estimating prevalence and / or incidence of a disease
- **Natural history of a disease**
  - exploring prognostic markers
- **Collect clinical data in context with a biobank**
- **Recruitment into RCTs**
  - assessing eligibility criteria
- …

AN ASIDE: RCT ELIGIBILITY CRITERIA BECOMING MORE COMPLEX OVER THE YEARS

- Example: systematic review of randomized placebo-controlled trials in relapsing multiple sclerosis (Steinvorth et al, 2013)



PURPOSES OF CLINICAL REGISTRIES

- **Observational data on treatments**
  - real-life treatment effects (population),
  - long-term follow-up (endpoint);
  - safety / pharmaco-epidemiology;
- **Comprehensive cohort design**
  - registry along side randomized controlled trial
  - patients not agreeing to randomization recruited into registry
- **Evidence synthesis**
  - combine data from a small RCT with observational data for confirmatory purposes in rare diseases / orphan indications

EXAMPLE: EUROPEAN REGISTER OF MULTIPLE SCLEROSIS (EUREMS)

EXAMPLE: EUREMS

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

MULTIPLE SCLEROSIS JOURNAL | MSJ

Research Paper

Multiple sclerosis registries in Europe – results of a systematic survey

Peter Flachenecker, Karoline Buckow, Maura Pugliatti, Vanja Bašić Kes, Mario A Battaglia, Alexey Boyko, Christian Confavreux[†], David Ellenberger, Danica Eskic, David Ford, Tim Friede, Jan Fuge, Anna Glaser, Jan Hillert, Edward Holloway, Eva Ioannidou, Ludwig Kappos, Elisabeth Kasilingam, Nils Koch-Henriksen, Jens Kuhle, Vito Lepore, Rod Middleton, Kjell-Morton Myhr, Anastasios Orologas, Susana Otero, Dorothea Pitschnau-Michel Otto Rienhoff, Jaume Sastre-Garriga, Tsveta Schyns-Liharska, Dragana Sutovic, Christoph Thalheim, Maria Trojano, Yan V Vlasov, Özgür Yaldizli, for the EUReMS Consortium[*]

Abstract

Background: Identification of MS registries and databases that are currently in use in Europe as well as a detailed knowledge of their content and structure is important in order to facilitate comprehensive analysis and comparison of data.

Methods: National MS registries or databases were identified by literature search, from the results of the MS Barometer 2011 and by asking 33 national MS societies. A standardized questionnaire was developed and sent to the registries' leaders, followed by telephone interviews with them.

Results: Twenty registries were identified, with 13 completing the questionnaire and seven being interviewed by telephone. These registries differed widely for objectives, structure, collected data, and for patients and centres included. Despite this heterogeneity, common objectives of the registries were epidemiology (n=10), long-term therapy outcome (n=8), healthcare research (n=9) and support/basis for clinical trials (n=8). While physician-based outcome measures (EDSS) are used in all registries, data from patients' perspectives were only collected in six registries.

Conclusions: The detailed information on a large number of national MS registries in Europe is a prerequisite to facilitating harmonized integration of existing data from MS registries and databases, as well as comprehensive analyses and comparison across European populations.

→ Heterogeneity

EXAMPLE: EU-CERT-ICD

http://www.eu-cert-icd.eu/

EU-CERT-ICD: SOURCES OF EVIDENCE

EU-CERT-ICD

- **sources**
  - publications
  - registries
  - cohort study
- **meta-analysis**
  - publication-based
  - individual-patient data (IPD)
  - combined: publications + IPD

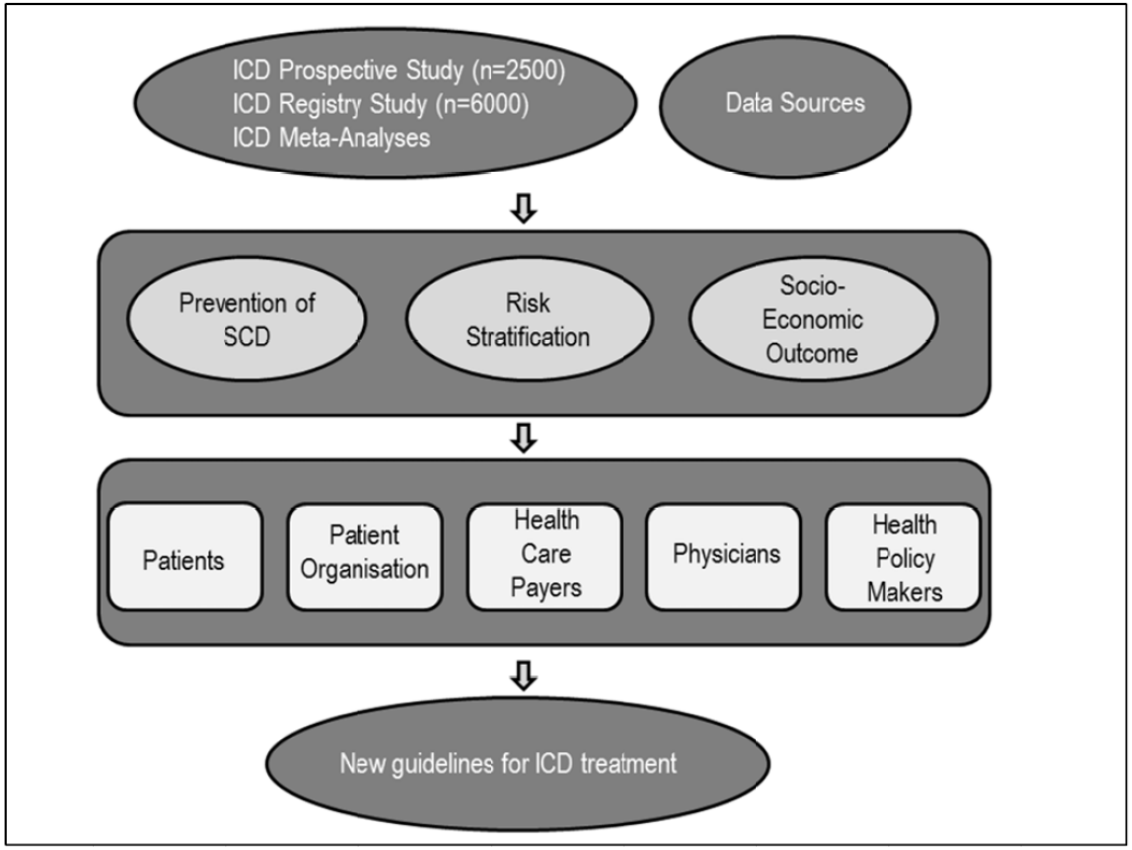


Figure 1: Concept of EU-CERT-ICD

EU-CERT-ICD REGISTRY

Home → The Project → Work Plan → WP02 Registry

EU-CERT-ICD

THE PROJECT THE G

Work Package 02: Retrospective ICD Registry

Work Package leader
Christian Sticherling

Universitätsspital Basel

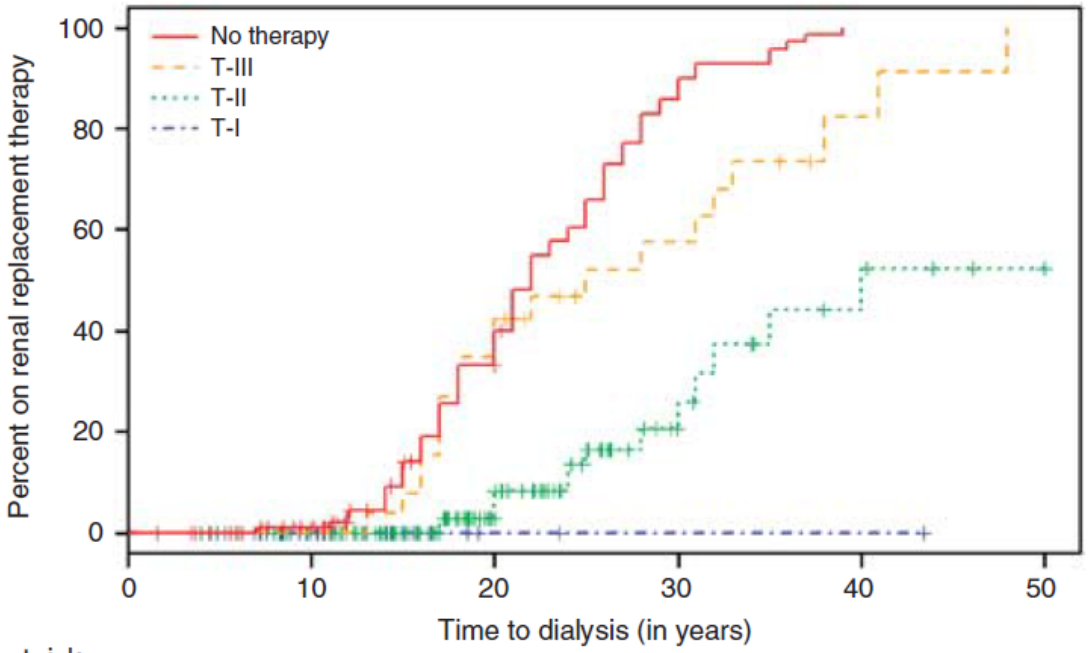
Objectives of the Work Package

- To establish a large central primary prophylactic ICD data base merging data from 11 partner institutions and preliminary statistical analyses
- To obtain blood samples for biomarker and genetic analyses in registry patients
- To exploit existing ECG recordings in registry patients

EXAMPLE: ALPORT DISEASE

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

- **Alport disease**
  - Rare genetic disease leading ultimately to kidney failure
- Data from the European registry suggest **ACE inhibition delays kidney failure** (Gross et al, 2012a)



No. at risk											
No therapy	109	105	96	75	50	29	10	5	0	0	0
T-III	26	26	26	25	17	10	8	5	2	1	0
T-II	115	113	105	84	52	31	15	9	7	4	3
T-I	33	32	20	8	2	1	1	1	1	0	0

EARLY PRO-TECT ALPORT TRIAL

- **Double-blind RCT in children**
  - Difficulties in recruitment to be expected
- **EARLY PRO-TECT Alport Trial** (Gross et al, 2012b)

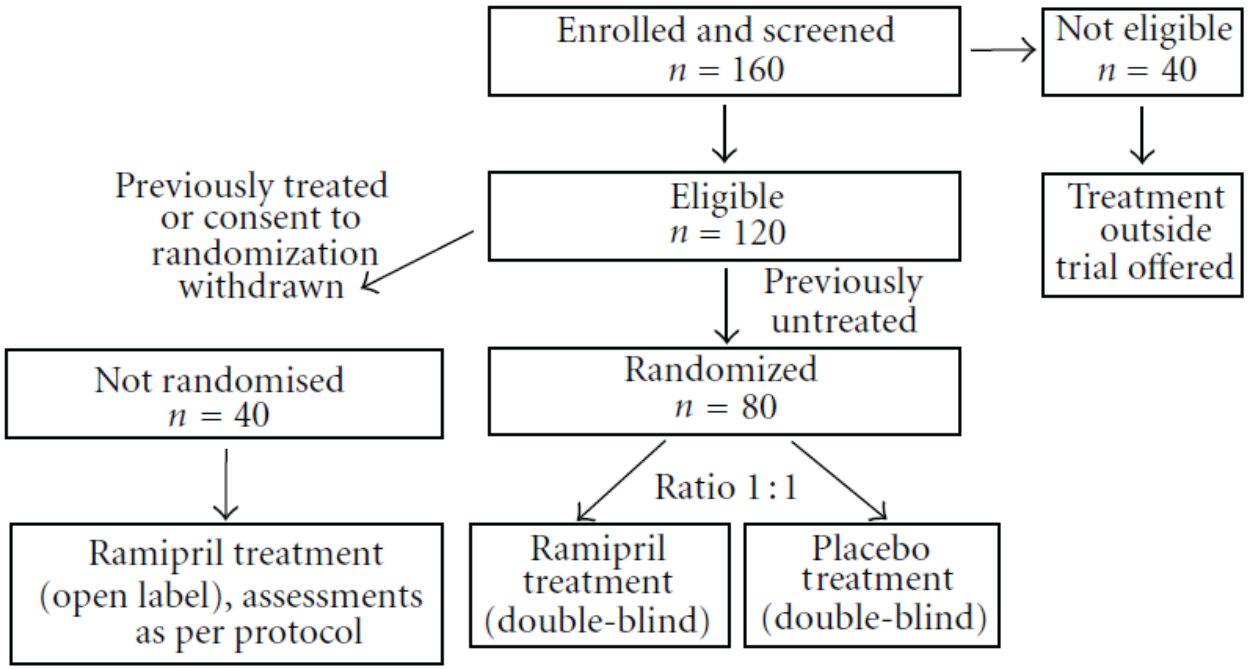


FIGURE 1: Study design of the EARLY PRO-TECT Alport trial.

REQUIREMENTS ON CLINICAL REGISTRY DEPENDING ON PURPOSE

- **Different requirements depending on specific purpose(s) of the clinical registry**
- **For example** …
  - Recruitment into RCTs: only basic information on demographics and disease course required
  - epidemiological registry to estimate prevalence / incidence: capture (nearly) all cases in a certain population
  - registry to study natural disease course / treatment effects: longitudinal data
  - registry to contribute to evidence synthesis with randomized controlled trial: registry needs to be sufficiently similar to RCT in terms of population and endpoints captured

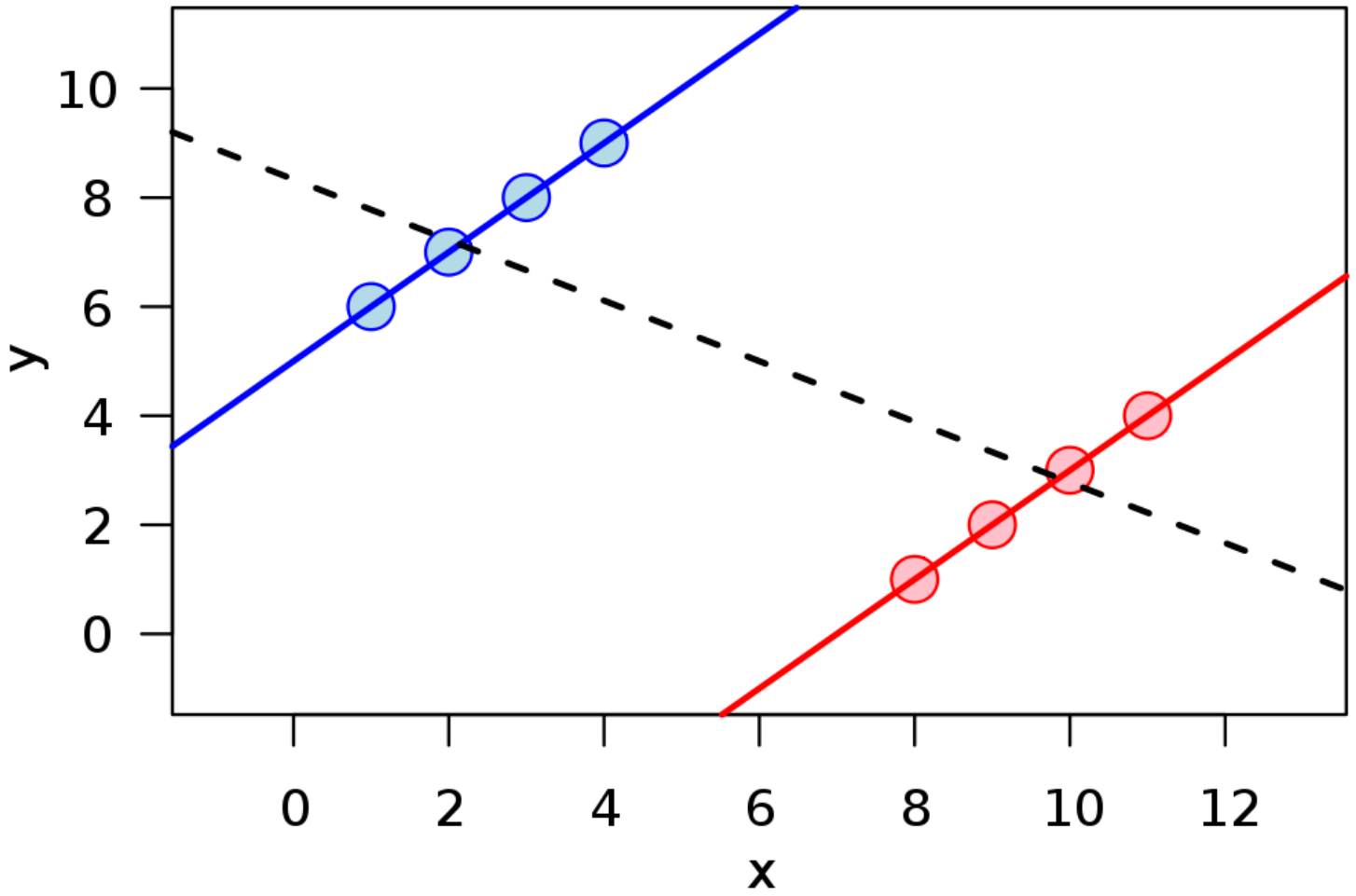
OUTLINE

- Background: Definition(s) and classification
- Purposes of clinical registries
- Statistical issues and methodologies
- Operational and logistical issues
- Conclusions and discussion

STATISTICAL ISSUES / METHODOLOGIES

- **Combining registries**
  - Pooling of data sets: Simpson's paradox
  - modelling between-registry heterogeneity
    - in particular when data are not collected under the same protocol, heterogeneity across registries expected
  - Statistical methods:
    - stratification by registry / centre
    - hierarchical models, individual patient data (IPD) meta-analysis (Debray et al (2015) Research Synthesis Meth)
  - Examples: EU-CERT-ICD

Simpson‘s Paradox



http://en.wikipedia.org/wiki/Simpson%27s_paradox

STATISTICAL ISSUES / METHODOLOGIES

- **Estimating treatment effects in observational data**
  - Problem of confounding in non-randomized treatment comparisons (selection problem)
  - Statistical methods: propensity scores (matching, stratifying, covariate, ...); …
  - Example from multiple sclerosis

EXAMPLE: MULTIPLE SCLEROSIS

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Association Between Use of Interferon Beta and Progression of Disability in Patients With Relapsing-Remitting Multiple Sclerosis

Afsaneh Shirani, MD
Yinshan Zhao, PhD
Mohammad Ehsanul Karim, MSc
Charity Evans, PhD
Elaine Kingwell, PhD
Mia L. van der Kop, MSc
Joel Oger, MD, FRCPC
Paul Gustafson, PhD
John Petkau, PhD
Helen Tremlett, PhD

Context Interferon beta is widely prescribed to treat multiple sclerosis (MS); however, its relationship with disability progression has yet to be established.

Objective To investigate the association between interferon beta exposure and disability progression in patients with relapsing-remitting MS.

Design, Setting, and Patients Retrospective cohort study based on prospectively collected data (1985-2008) from British Columbia, Canada. Patients with relapsing-remitting MS treated with interferon beta (n=868) were compared with untreated contemporary (n=829) and historical (n=959) cohorts.

Main Outcome Measures The main outcome measure was time from interferon beta treatment eligibility (baseline) to a confirmed and sustained score of 6 (requiring a cane to walk 100 m; confirmed at >150 days with no measurable improvement) on the Expanded Disability Status Scale (EDSS) (range, 0-10, with higher scores indicating higher disability). A multivariable Cox regression model with interferon beta treatment included as a time-varying covariate was used to assess the hazard of disease progression associated with interferon beta treatment. Analyses also included propensity score adjustment to address confounding by indication.

Results The median active follow-up times (first to last EDSS measurement) were as follows: for the interferon beta–treated cohort, 5.1 years (interquartile range [IQR], 3.0-7.0 years); for the contemporary control cohort, 4.0 years (IQR, 2.1-6.4 years); and for the historical control cohort, 10.8 years (IQR, 6.3-14.7 years). The observed outcome rates for reaching a sustained EDSS score of 6 were 10.8%, 5.3%, and 23.1% in the 3 cohorts, respectively. After adjustment for potential baseline confounders (sex, age, disease duration, and EDSS score), exposure to interferon beta was not associated with a statistically significant difference in the hazard of reaching an EDSS score of 6 when either the contemporary control cohort (hazard ratio, 1.30; 95% CI, 0.92-1.83; $P=.14$) or the historical control cohort (hazard ratio, 0.77; 95% CI, 0.58-1.02; $P=.07$) were considered. Further adjustment for comorbidities and socioeconomic status, where possible, did not change interpretations, and propensity score adjustment did not substantially change the results.

Conclusion Among patients with relapsing-remitting MS, administration of interferon beta was not associated with a reduction in progression of disability.

JAMA. 2012;308(3):247-256 www.jama.com

MULTIPLE SCLEROSIS (MS) IS a chronic disease that often affects people in the prime of their lives. A key feature of MS is clinical progression of the disease over time manifested by the accumulation of disability. Interferon beta drugs are the most widely prescribed disease-modifying drugs (DMDs) approved by the US Food and Drug Administration for the treatment of relapsing-onset MS, the most common MS disease course. Although a substantial reduction in brain lesion development, as evidenced by magnetic resonance imaging (MRI),[1] and a one-third relative reduction in relapse frequency were demonstrated in the

EXAMPLE: MULTIPLE SCLEROSIS

In an editorial Derfuss and Kappos comment:

- "Does this mean that in the "real world" and with longer follow-up, the benefits of interferon beta demonstrated in controlled trials are no longer relevant and that administration of interferon beta should not be prescribed and reimbursed?"
- "**Lacking evidence of treatment effect is not proof of lacking effect**."
- "Furthermore, although methodologically sound, this study cannot avoid the inherent challenges of data analysis and interpretation in nonrandomized observational studies. **Sophisticated statistical methods may help adjust for known unequally distributed baseline variables but cannot account for subtle unmeasured selection criteria as sources of bias**."

STATISTICAL ISSUES / METHODOLOGIES

- **Missing data**
  - Missing data can occur for different reasons: e.g. lower standards in data capturing than in RCT; different centres collect data under (slightly) different protocols
  - Statistical methods: a variety of methods available
  - Example: Predicting survival in heart failure

EXAMPLE: HEART FAILURE

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

European Heart Journal (2013) **34**, 1404–1413
doi:10.1093/eurheartj/ehs337

CLINICAL RESEARCH
Chronic heart failure

Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies

Stuart J. Pocock[1]*, Cono A. Ariti[1], John J.V. McMurray[2], Aldo Maggioni[3], Lars Køber[4], Iain B. Squire[5], Karl Swedberg[6], Joanna Dobson[1], Katrina K. Poppe[7], Gillian A. Whalley[7], and Rob N. Doughty[7], on behalf of the Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC)

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK; [2]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK; [3]ANMCO Research Centre, Florence, Italy; [4]Rigshospitalet—Copenhagen University Hospital, Copenhagen, Denmark; [5]Department of Cardiovascular Sciences, The University of Leicester, Leicester, UK; [6]Sahlgrenska University, Hospital/Östra, Göteborg, Sweden; and [7]Department of Medicine, University of Auckland, Auckland, New Zealand

Received 22 May 2012; revised 3 August 2012; accepted 13 September 2012; online publish-ahead-of-print 24 October 2012

See page 1391 for the editorial comment on this article (doi:10.1093/eurheartj/ehs363)

Aims Using a large international database from multiple cohort studies, the aim is to create a generalizable easily used risk score for mortality in patients with heart failure (HF).

Methods and results The MAGGIC meta-analysis includes individual data on 39 372 patients with HF, both reduced and preserved left-ventricular ejection fraction (EF), from 30 cohort studies, six of which were clinical trials. 40.2% of patients died during a median follow-up of 2.5 years. Using multivariable piecewise Poisson regression methods with stepwise variable selection, a final model included 13 highly significant independent predictors of mortality in the following order of predictive strength: age, lower EF, NYHA class, serum creatinine, diabetes, not prescribed beta-blocker, lower systolic BP, lower body mass, time since diagnosis, current smoker, chronic obstructive pulmonary disease, male gender, and not prescribed ACE-inhibitor or angiotensin-receptor blockers. In preserved EF, age was more predictive and systolic BP was less predictive of mortality than in reduced EF. Conversion into an easy-to-use integer risk score identified a very marked gradient in risk, with 3-year mortality rates of 10 and 70% in the bottom quintile and top decile of risk, respectively.

Conclusion In patients with HF of both reduced and preserved EF, the influences of readily available predictors of mortality can be quantified in an integer score accessible by an easy-to-use website www.heartfailurerisk.org. The score has the potential for widespread implementation in a clinical setting.

EXAMPLE: HEART FAILURE

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

- **Summary table to describe extent of missing data**

Table 3 Extent of missing data

Model variable	Studies with no data		Studies with some data		Total patients missing data
	Studies	Missing patients	Studies	Missing patients	
Age	0	0	0	0	0
Gender	0	0	0	0	0
BMI	17	14 515	13	2686	17 201
Current smoker	6	9166	24	448	9614
SBP	9	12 016	21	276	12 292
Diabetes	1	348	29	341	689
NYHA class	5	2503	25	1128	3631
Ejection fraction	6	3279	24	3558	6837
COPD	10	16 788	20	253	17 041
HF duration	20	11 679	10	1066	12 745
Creatinine	5	2800	25	17 245	20 045
Beta-blocker	3	7890	27	709	8599
ACE-I/ARB	1	97	29	649	746

BMI, body mass index; SBP, systolic blood pressure; NYHA, New York Heart Association; COPD, chronic obstructive pulmonary disease; HF, heart failure; ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin-receptor blockers.

- **Methods**: “Missing values are handled by multiple imputations using chained equations.”
- **References**: White & Royston (2009); White et al. (2011)

Statistics in Medicine

Research Article

Received 23 December 2014, Accepted 17 November 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6837

Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates

M. Quartagno[a*†] and J. R. Carpenter[a,b]

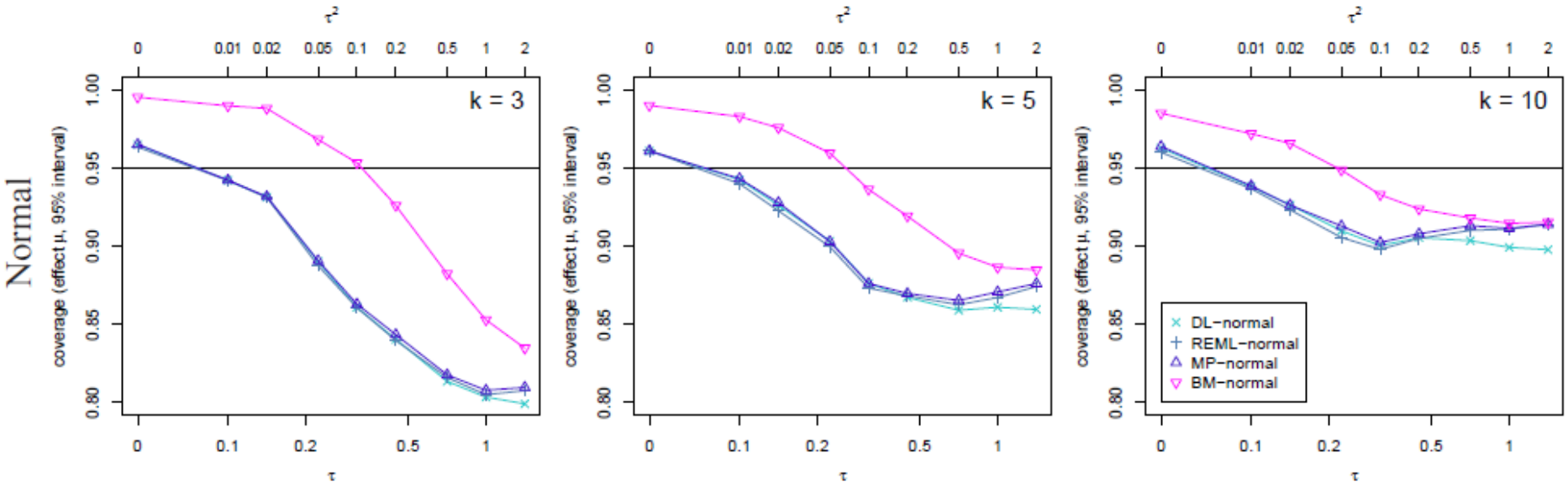
Recently, multiple imputation has been proposed as a tool for individual patient data meta-analysis with sporadically missing observations, and it has been suggested that within-study imputation is usually preferable. However, such within study imputation cannot handle variables that are completely missing within studies. Further, if some of the contributing studies are relatively small, it may be appropriate to share information across studies when imputing. In this paper, we develop and evaluate a joint modelling approach to multiple imputation of individual patient data in meta-analysis, with an across-study probability distribution for the study specific covariance matrices. This retains the flexibility to allow for between-study heterogeneity when imputing while allowing (i) sharing information on the covariance matrix across studies when this is appropriate, and (ii) imputing variables that are wholly missing from studies. Simulation results show both equivalent performance to the within-study imputation approach where this is valid, and good results in more general, practically relevant, scenarios with studies of very different sizes, non-negligible between-study heterogeneity and wholly missing variables. We illustrate our approach using data from an individual patient data meta-analysis of hypertension trials. © 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

STATISTICAL ISSUES / METHODOLOGIES

- **Calendar time effects**
  - Statistical methods: e.g. smooth (flexible parametric) models, change-point methods
  - Example: Registries in hip replacement (Friede & Henderson (2003) Stat Med)
- **Evidence synthesis**
  - Statistical methods: hierarchical models; power priors; recent overview provided by Viele et al (2014) Pharm Stat
  - Modelling of heterogeneity important, but estimation of challenging with only few studies (a situation frequently encountered)
  - Example from Alport disease (rare disease)

ESTIMATION OF BETWEEN-TRIAL HETEROGENEITY

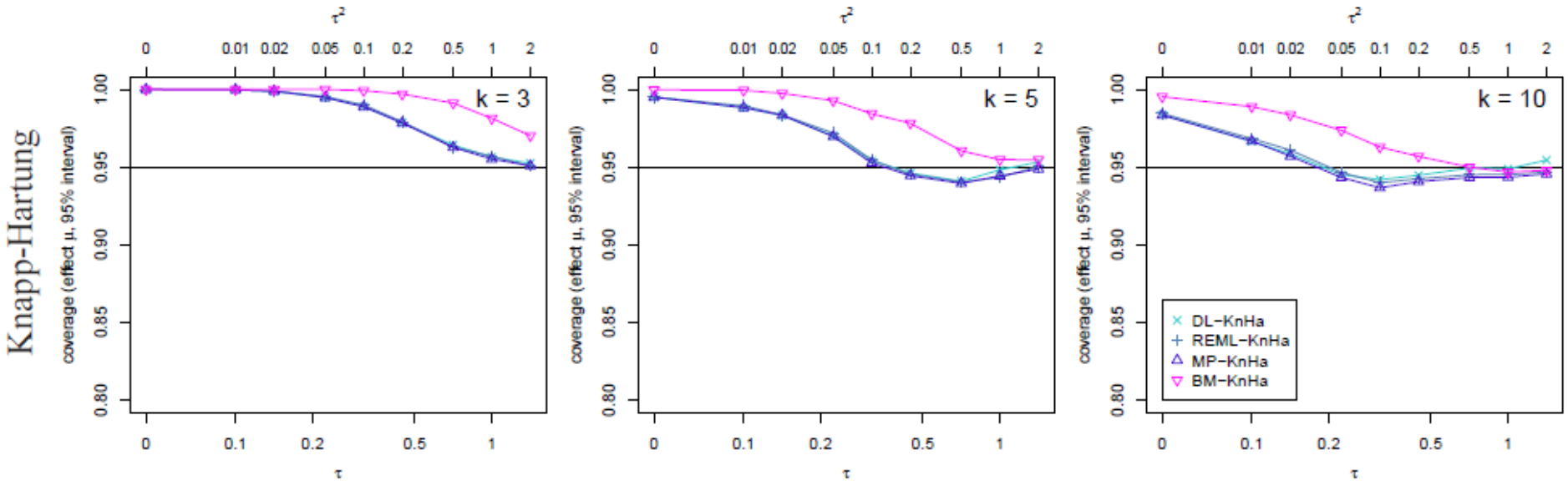
- Coverage probability for confidence intervals of combined effect
- Construction of confidence intervals using normal quantiles
- Estimators: DerSimonian-Laird (DL), restricted maximum likelihood (REML), Mandel-Paule (MP), **Bayes-modal** (BM)



Friede et al. (2015)

ESTIMATION OF BETWEEN-TRIAL HETEROGENEITY

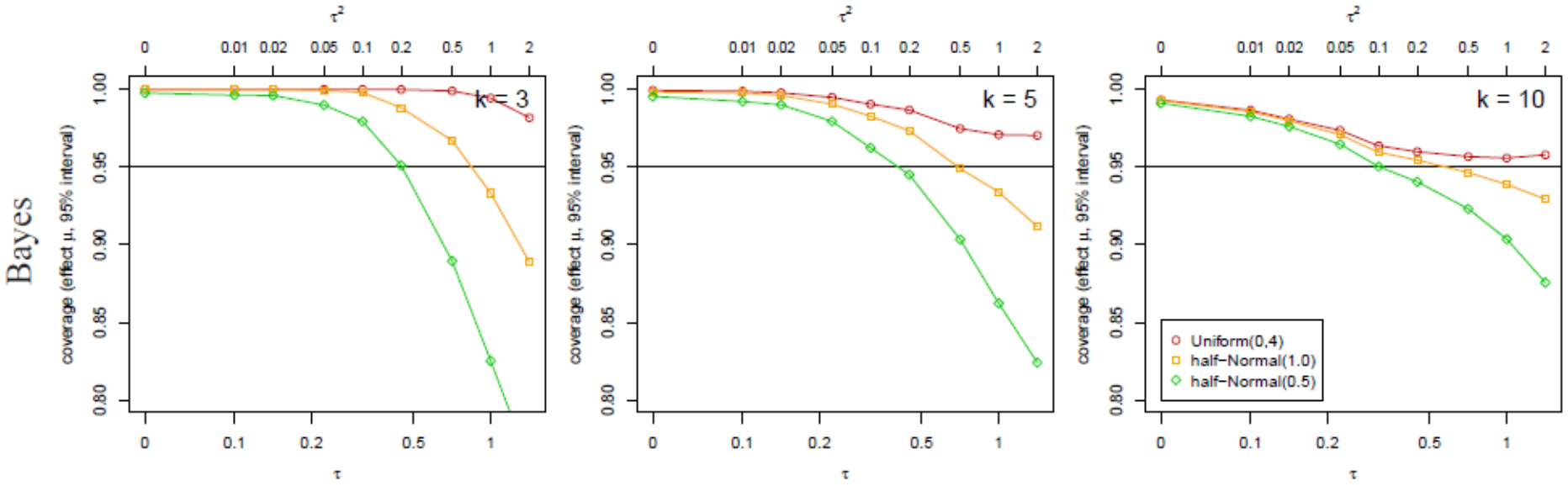
- Coverage probability for confidence intervals of combined effect
- Construction of confidence intervals using **Knapp-Hartung method** (using t-quantiles and scaling of standard error)



Friede et al. (2015)

ESTIMATION OF BETWEEN-TRIAL HETEROGENEITY

- Coverage probability for credibility intervals of combined effect
- Bayes with “weakly informative” priors for tau



Friede et al. (2015)

STATISTICAL ISSUES / METHODOLOGIES

- **Comprehensive cohort design**
  - Statistical methods: see Schmoor et al (1996)
  - Comparisons between randomized and non-randomized patients: (a) baseline characteristics, (b) outcome (e.g. survival), and (c) treatment effect
  - Example: DZHK VAD study in patients awaiting HTx

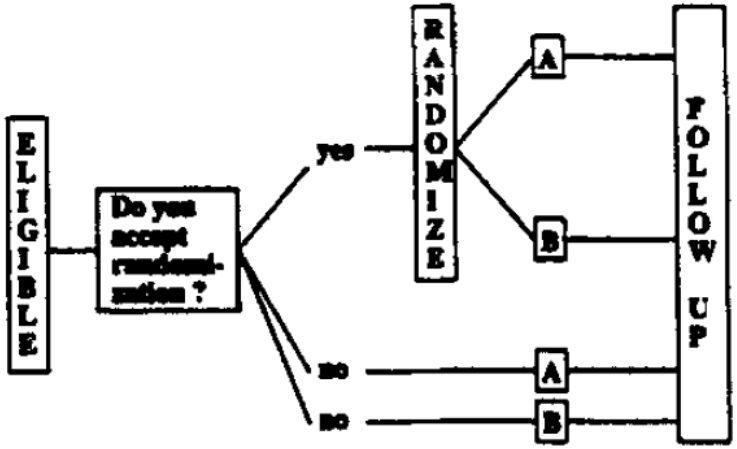


Figure 1. Design of the Comprehensive Cohort Study

OUTLINE

- Background: Definition(s) and classification
- Purposes of clinical registries
- Statistical issues and methodologies
- Operational and logistical issues
- Conclusions and discussion

OPERATIONAL / LOGISTICAL ISSUES

- **Data structure**
  - minimal data set, modular system
- **Ethical / legal aspects**
  - multi-national registries
- **Ownership / organization**
  - Academic institutions
  - Clinical community
  - Patient organisations
  - Companies (pharma / CRO) ...

OPERATIONAL / LOGISTICAL ISSUES

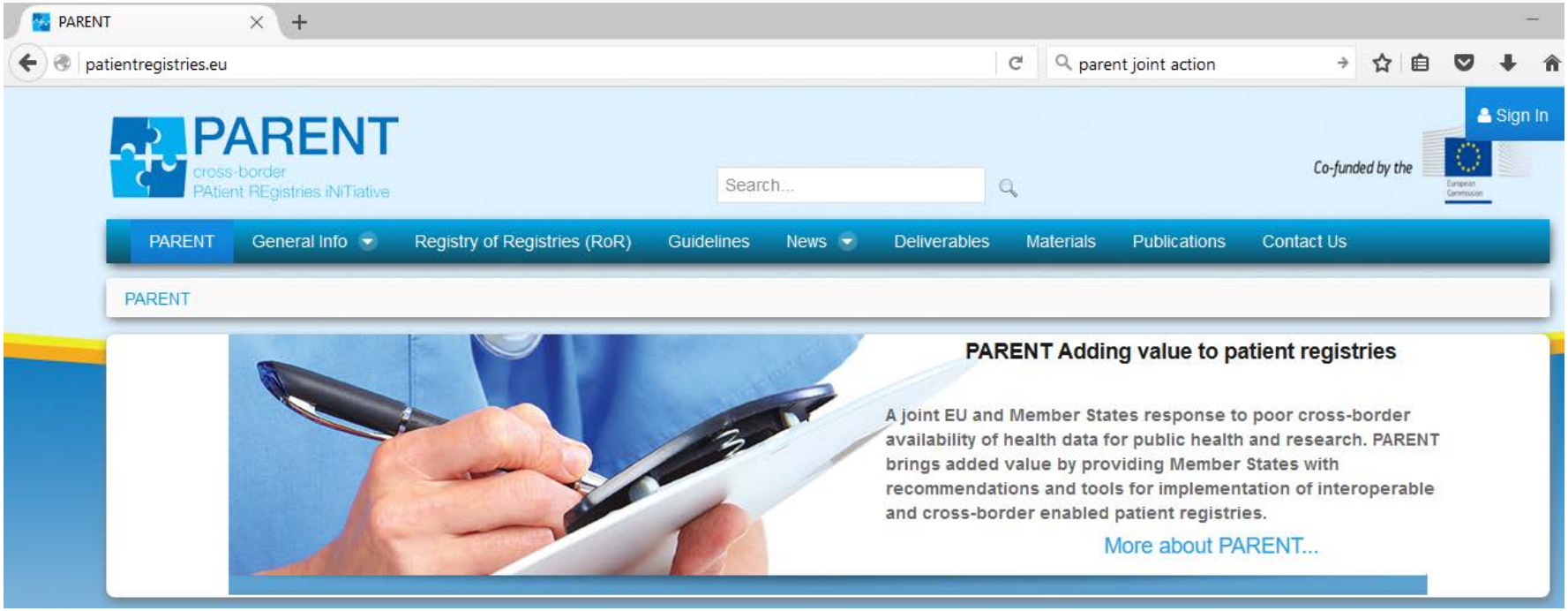
- **Use of and access to data**
- **Sustainability**
  - Funding
  - Content development
  - Technical requirements
- **Linking registries with other sources**
  - Biobanks
  - Imaging repositories
  - Patient reported outcomes (entered directly by patients)

OUTLINE

- Background: Definition(s) and classification
- Purposes of clinical registries
- Statistical issues and methodologies
- Operational and logistical issues
- Conclusions and discussion

RESOURCES – A FEW POINTERS

- **US Agency for Healthcare Research and Quality (AHQR)** publication (available online): Registries for Evaluating Patient Outcomes: A User's Guide
- **cross-border PAtient REgistries iNiTiative (PARENT)**



CONCLUSIONS AND DISCUSSION

- Clinical registries useful tool to supplement our tool box in clinical research
- Requirements on a registry depend on its purpose
- Use of clinical registries in confirmatory sense depending on setting (e.g. rare disease, devices, …)
- Sustainability appears to be a big hurdle in many settings

ACKNOWLEDGEMENTS

- "Innovative methodology for small populations research" (InSPiRe) received funding from the EU's 7th Framework Programme for research, technological development and demonstration under grant agreement n° FP HEALTH 2013 – 602144
- EU-CERT-ICD is funded by the European Commission within the 7th Framework Programme under Grant Agreement n° 602299
- EUReMS received co-funding from the EU in the framework of the Second Health Programme 2008–2013, Priority Area: 3.3.2, Action 3.3.2.7
- EARLY PRO-TECT ALPORT TRIAL received funding from the BMBF (German Ministry of Education and Research)