Searching for the ideal clinical study design:

The Quest for the Holy Grail?

Emmanuel Lesaffre

I-Biostat, KU Leuven, Leuven, Belgium

UHASSELT
I-BioStat
KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

CONFLICT OF INTEREST DISCLOSURE

I have **no** potential conflict of interest to report

Contents

- Aims of clinical research
- Specifics of geriatric population
- Classical epidemiological study designs: **theory**
- Classical epidemiological study designs: **practice**
- Practical conclusions

Focus is on comparison of drug treatments
but the talk
also applies to other interventions

Aims of clinical research

- **Aims of clinical research are**:
  - **In general**: establish/evaluate risk factors for diseases and symptoms
  - **Here**: selecting the best treatment
  - **Also**: determine **which** patient should receive **what** treatment

1 million $ question:

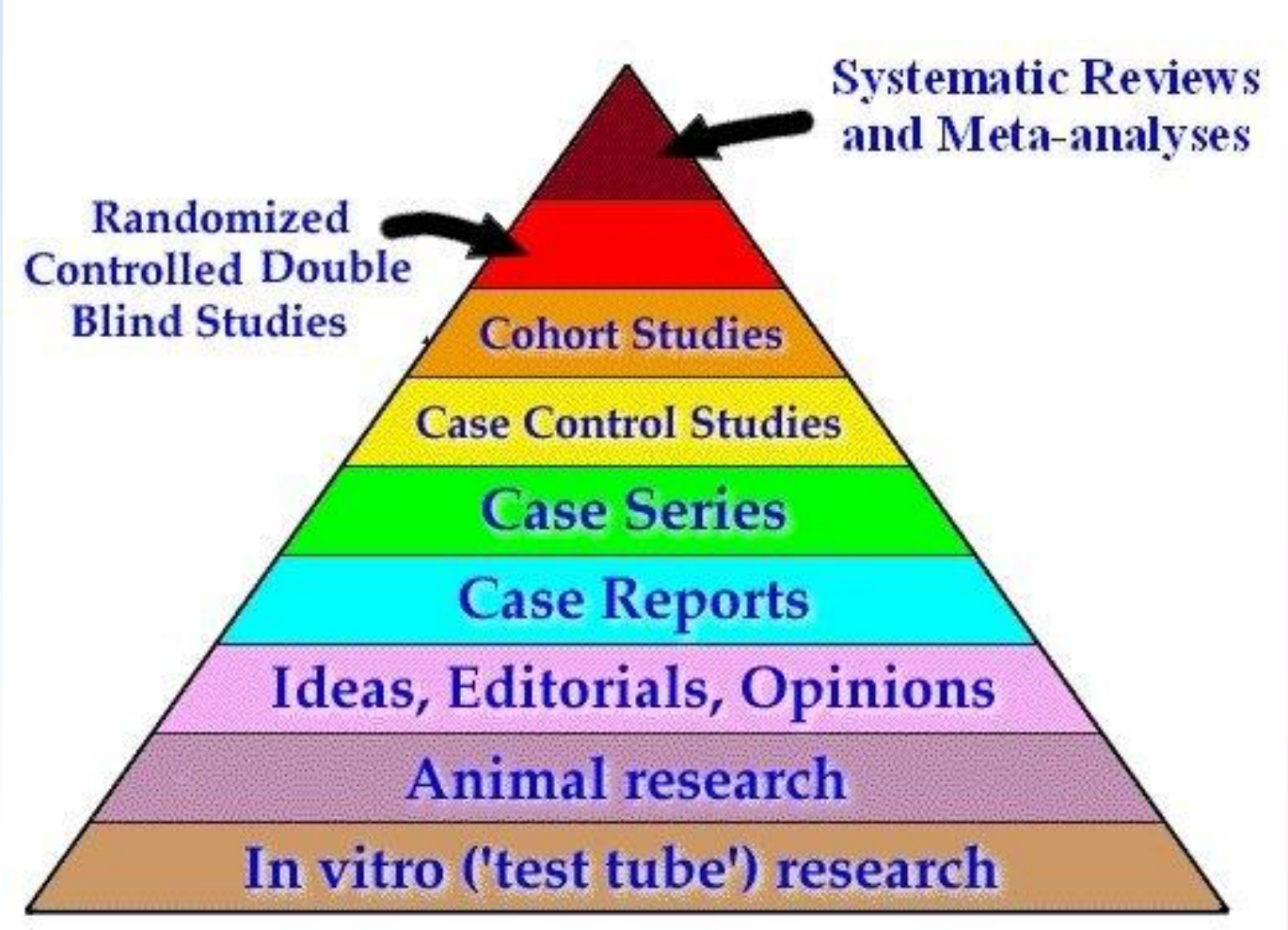
Which study design to answer these questions?

Specifics of **geriatric** population

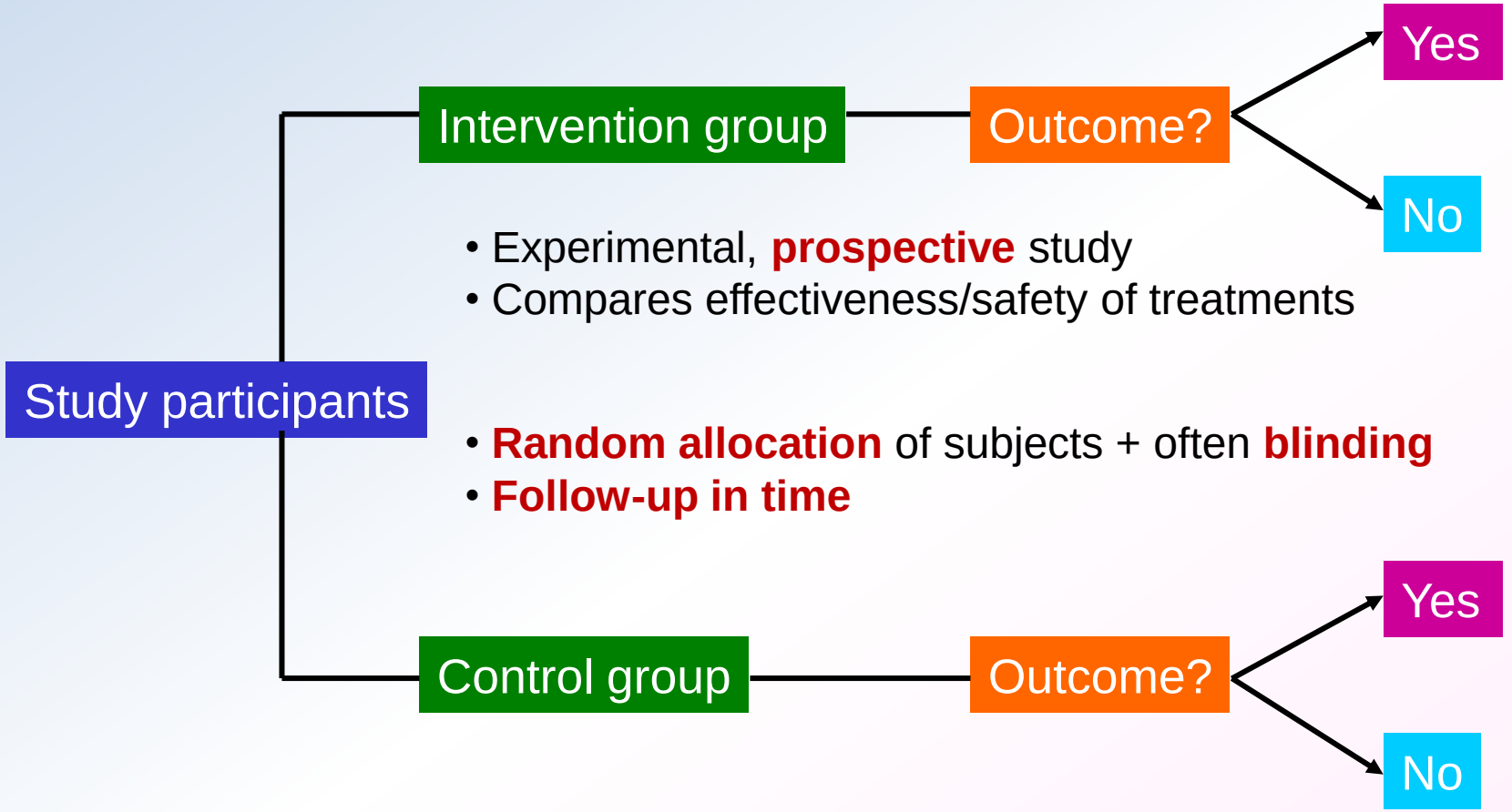
- Multiple comorbidities
- Many concomitant medications
- Higher number of dropouts due to death
- Age range restrictions in RCTs
- …

Classical clinical study designs: **theory**

Pyramid of evidence



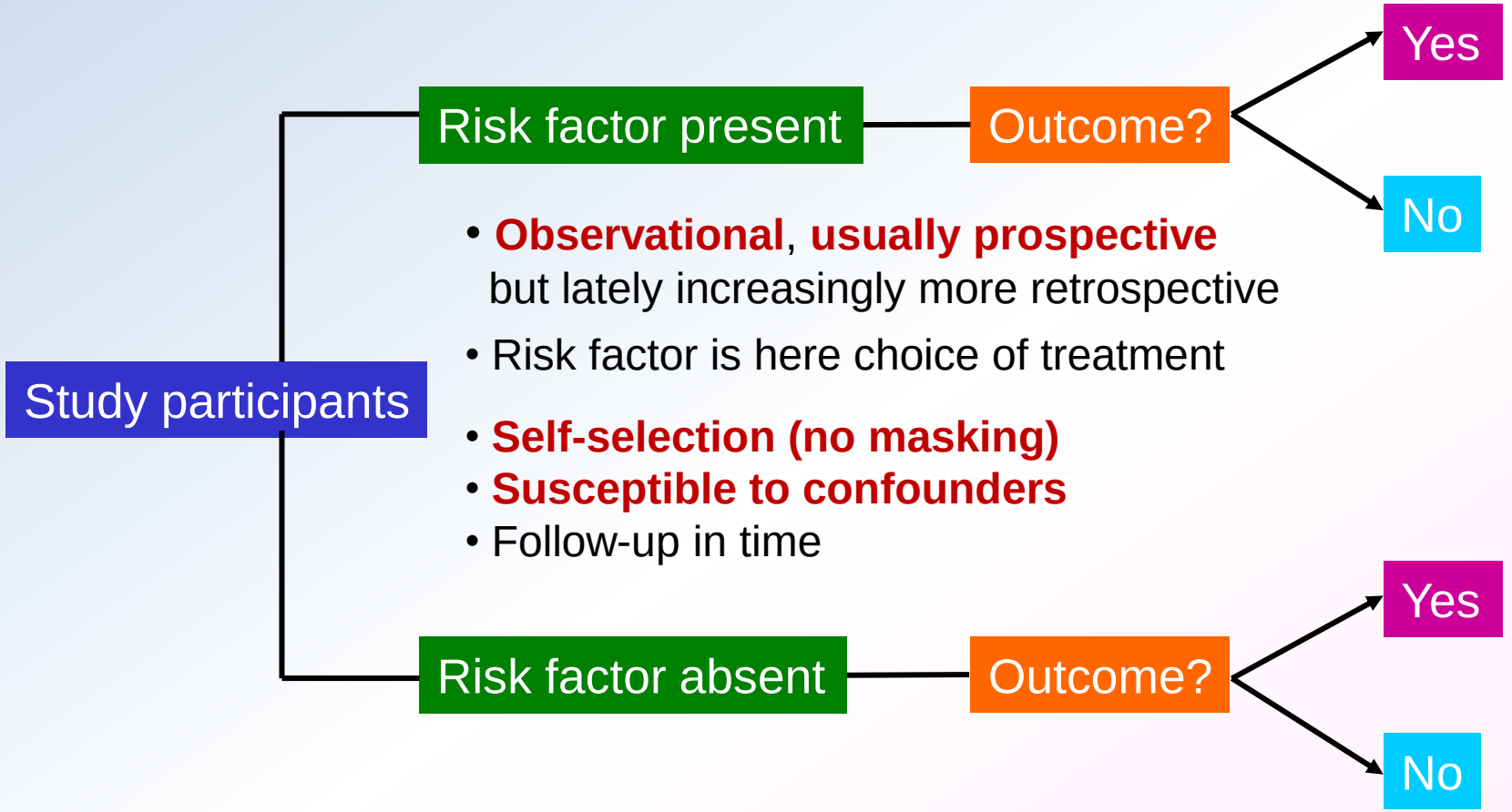
Randomized Controlled Trial



- **Costly and time consuming**, but low potential for bias
- High level of evidence: **allows for causal claims, if done properly**

Cohort study design

Longitudinal observational study, real life study, ...



- **Observational**, **usually prospective** but lately increasingly more retrospective
- Risk factor is here choice of treatment
- **Self-selection (no masking)**
- **Susceptible to confounders**
- Follow-up in time

- **Time-consuming**, loss to follow-up often a problem
- High level of evidence, but **only association** can be measured

Classical clinical study designs: **practice**

Randomized Controlled Trial (RCT)

- RCT: **gold standard** for clinical research, **at least in theory**
- But theory is often different from practice
- Evaluation in practice:
  - Quality of data
  - Statistical aspects (internal validity) & causality ⇔ association
  - What is measured?
  - External validity
  - Efficacy ⇔ safety

Focus on comparison of 2 treatments for efficacy but also safety

RCT: quality of data

- RCTs are prospective
- Patients are monitored, which implies:
  - Quality of data is better than for retrospective studies
  - Less missing data than for retrospective studies
  - Quality of data also (often) better than for real life studies
  - Less misclassified symptoms, comorbidities, …

RCT: statistical aspects

- **Randomisation:** treatment groups are **balanced at start** for all known and **UNKNOWN** confounding factors
- **Blinding:** disentangles psychological and biological effect
- **Statistical implications:**
  - **No statistical comparison, no P-values at baseline!**
  - **Simple statistical tests** can be used: t-test, χ^2-test, …
- But, only when one takes into account appropriately:
  - Missing data, dropouts, …
  - Protocol violators, compliance, …
- **RCT** is the ONLY design that allows to establish **causal** relationship: measured effect of treatment is **only** due to treatment

RCT: what is measured?

- **Exclusion criteria in RCTs** imply
  - Patients with selected comorbidities are not included
  - Patients taking certain concomitant treatments are not allowed
- Patients in RCTs are **closely monitored**

$\Rightarrow$ **Upper bound** of treatment effect is measured in RCTs

RCT: external validity

- **Exclusion criteria in RCTs** imply
  - The selected patients are **not representative** for the total patient population of interest (**selection bias**)
  - That is, **external validity of RCTs is often low**
- **Geriatric studies** generally suffer even more from exclusion criteria
  - Age limits
  - Avoiding comorbidities
  - Restricting concomitant medication

Underrepresentation of elderly in RCTs

Including older people in clinical research

Benefits shown in trials in younger people may not apply to older people

BMJ VOLUME 331 5 NOVEMBER 2005 bmj.com

Perspectives in Clinical Research

Home
Current issue
Instructions
Submit article

Perspect Clin Res. 2015 Oct-Dec; 6(4): 184–189.
doi: 10.4103/2229-3485.167099

PMCID: PMC4640010

Elderly patients' participation in clinical trials

Premnath Shenoy and Anand Harugeri

Abstract

The elderly population is a large and the fastest-growing portion of the population worldwide. The elderly make up the lion's share of patients for certain health conditions including cancer, cardiovascular disease, arthritis, and Parkinson's disease, among others in most parts of the world. Furthermore, elderly make up the majority of patients for many medications treating chronic conditions. Typically, clinical trials conducted in adult population include patients between the ages of 18 and 64 years. However, drugs should be studied in all age groups and trial participants should be representative of the patient population receiving the therapy in daily medical practice. Elderly patients are poorly represented in clinical trials.

RCT: efficacy ⇔ safety

- **Same principles apply to safety as to efficacy**
- But, RCTs are designed to detect treatment effects (efficacy)
- RCTs are (most) often **underpowered** to evaluate safety:
  - **Rare** adverse events cannot be detected with realistic study sizes
  - Some adverse events only occur after long periods of drug intake

Hyperkalemia ⇔ spironolactone treatment

Juurlink et al. (NEJM, 2004)

- **RALES study** (1999): **spironolactone** significantly improves outcomes (symptoms heart failure, 30% reduction in mortality) in patients with **severe heart failure**.
- But: **ACE inhibitors** are also indicated in these patients
- Spironolactone **can provoke life-threatening hyperkalemia** when combined with ACE inhibitors
- In RALES study no strong evidence for such a dangerous effect was found, but **"Clinical trial setting and actual practice are particularly relevant for older patients, most of whom would not have been included in RALES."**
- A **population-based time-series study** (registry in Ontario): 1,6 million adults > 66 years, period: 1994 - 2001
- **Result 1**: **significant relation** ($P < 0.001$) between subscription of spironolactone and hospitalization for hyperkalemia/heart failure from 34/1000 to 149/1000
- **Result 2**: Mortality increased from 0.3/1000 to 2.0/1000 ($P<0.001$)

Longitudinal observational/real life studies

- Of the 3 classical epidemiological designs (cohort, case-control, cross-sectional) the **cohort design is by far best** to establish an association between risk factors and the occurrence of diseases/symptoms
- Cohort/longitudinal/real life data can be obtained from:
  - Phase IV studies
  - (Longitudinal) registries
  - …
- **What is gained/lost compared to a RCT?**

Cohort design ⇔ RCT

- **Data quality:** cohort designs are often prospective ⇒ data quality data better than for CC & X-sectional studies, but less than for RCTs
- **Statistical aspects:** since there is self-selection and no masking, the **statistical procedures are more complicated**, see next slides
- **Causality ⇔ association:** only association can be shown, although sophisticated statistical procedures try to come close to a RCT
- **What is measured:** the effect and safety of treatments in **real life** settings, but often the comparison is not (adequately) controlled
- **External validity:** **highly relevant** to the general population, but the message is **not always clear**
- **Safety:** real life studies are typically done over longer periods with many more patients, hence **better powered to find rare AEs**

Cohort design: statistical aspects

- **Self-selection:** treatment groups are imbalanced at baseline
- **How to correct for imbalance?**
  - **Perfect correction is NOT possible**
  - "Multivariate" analyses (logistic & Cox regression) are performed to correct for imbalances
  - Nowadays, **propensity score analyses** are popular
  - One could also **match** the patients in the two treatment groups
  - But **one can never correct for not-observed** imbalances
  - **In addition**: one is never sure that the statistical model is correct!

Cohort design: propensity score analysis

- **Univariate analysis:** 2 treatment groups with respect to outcome
- **"Multivariate" analysis:** Correct for important observed covariates with logistic regression, Cox regression, …
- **Propensity score analysis: aims to mimic a RCT**
  1. Take many covariates (even those that do not have a relationship with outcome)
  2. Predict the treatment group (using logistic regression) from all those covariates
  3. Obtain the score to predict allocation to one treatment (= probability to choose that treatment)
  4. Apply logistic/Cox regression with propensity score + other important covariates to predict outcome
  5. Possibly apply stratification or matching instead

Risk of Lower and Upper Gastrointestinal Bleeding, Transfusions, and Hospitalizations With Complex Antithrombotic Therapy in Elderly Patients

Neena S. Abraham, MD, MSCE; Christine Hartman, PhD; Peter Richardson, PhD; Diana Castillo, BA; Richard L. Street, Jr, PhD; Aanand D. Naik, MD

Background—Complex antithrombotic therapy (CAT) prescribed to elderly patients increases the risk of gastrointestinal bleeding. We quantified upper (UGIE) and lower gastrointestinal (LGIE) events, transfusions, and hospitalizations in a national cohort of elderly veterans prescribed CAT.

Conclusions—Among elderly patients, CAT-related LGIE and UGIE are clinically relevant risks resulting in increased hospitalizations and transfusions. **(*Circulation*. 2013:128:1869-1877.)**

The burden of CAT-related GI bleeding remains largely unknown among elderly patients, an emerging population of patients systematically excluded from pivotal CAT trials. The best evidence would come from a double-blind, randomized, controlled trial with multiple arms that assesses GI bleeding as the outcome of interest. However, this would be unethical, given the known GI bleeding risk factor of advanced age.[3] To quantify the risk of CAT-related GI bleeding, we examined UGIE and LGIE risk and associated transfusions and hospitalizations in a national cohort of elderly patients. Our goal was to quantify the real-life number needed to harm (NNH) associated with CAT-related GI bleeding and to measure 2 patient-centered outcomes (ie, need for transfusion and bleeding requiring hospitalization) in a national cohort of elderly pa

A retrospective cohort study on veterans (60-99 yrs)

Potential Risk Factors and Confounding Variables

Pharmacological risk factors, including prescription of nonsteroidal anti-inflammatory drugs (NSAIDS), both traditional and cyclooxygenase-2 selective, steroids, selective serotonin reuptake inhibitors (SSRIs) and selective norepinephrine reuptake inhibitors, low-molecular-weight heparin, and heparin overlapping with CAT prescription, were identified (Table II in the online-only Data Supplement). We

PPI = proton pump inhibitor
TRIP = anticoagulant-antiplatelet-ASA

tion (using our previously validated algorithm),[9] and a time-dependent Elixhauser comorbidity Index score[9] (Table 1). Unadjusted analyses were used to assess patient characteristics that might be associated with a greater likelihood of being prescribed a particular prescription strategy (ie, confounding by indication); prescription of a PPI or triple therapy were associated with unique patient characteristics that influenced outcomes in unadjusted analyses. Thus, a propensity score[10] was calculated to estimate the conditional probability of a patient being prescribed a PPI or TRIP. Survival analyses were then stratified by propensity score quintiles to adjust for bias when estimating treatment effects,[11] and treatment effect was assessed within each stratum

NOACs ⇔ warfarin

- **Question:** What is value of "real-life" studies?
- **Setting:** Patients suffering from atrial fibrillation

1. Up to recently warfarin was standard treatment for stroke prevention
2. Four **N**on-vitamin K antagonist **O**ral **A**nti**C**oagulants (NOACs) have shown in RCTs to be non-inferior to warfarin, with apixaban superior to warfarin for the primary outcome but also for bleeding
3. No head-to-head RCT has been set up, but several "**real-life**" studies have been organized to compare GI bleeding incidence
4. All studies make use of "multivariate analyses" and many also include (two types of) **propensity score analyses**
5. **Results**: superiority of apixaban wrt GI bleeding compared to warfarin confirmed in "real-life" analysis & about same results for other NOACs

Real World Evidence

- **Value of RCT**
  - Proof of biological effect of treatment in 'ideal' situation (upper bound?)
  - But, obtained on a non-randomly selected and small set of patients
- **Value of real life studies**
  - Evidence is needed of how treatments work in real life
  - But, in general there is no assurance that quality of data in observational studies is good enough
  - Recently, there is much interest in how to combine information from:
    - Electronic data bases
    - Phase IV studies
  - And to examine on how to increase quality of data
  - But there is definitely a need to complement the information obtained from RCTs, **for a better personalized medicine**

It is now time for more interesting talks

Back up slides

Practical conclusions

- **What about case-control & cross-sectional studies?**
  - Prone to many more biases
  - Chicken-egg problem

- **Points to consider**
  - Amount and type of missing data, dropouts
  - Dropout due to death is different from dropout due to lack of efficacy, safety issues, …
  - Additional studies beyond RCTs are needed for geriatric population because of in- and exclusion criteria
  - Meta-analyses of subgroups of elderly patients in RCTs is still an option