Patient-based clinical trials

PD Dr. Pablo E. Verde

Heinrich Heine University Düsseldorf

ESB2023-Maastricht

pabloemilio.verde@hhu.de

09 July 2023

Overview of the lecture

- Introduction to regulatory aspects in clinical trials
- Part 1: Questions before the trial starts
  - Planning the trial, sample size and dealing with uncertainty
  - Randomization, inference and trial populations
- Part 2: Questions while performing the trial
  - Patient recruitment issues
  - Sequential designs and other issues
- Part 3: Questions after the data collection
  - Reporting and statistical software
  - Issues with outliers and missing data
- Part 4: In silico clinical trials and medical algorithms
  - Evaluation of the impact of a medical algorithm
  - Bayesian meta-analysis of medical decision tools

Introduction:

Regulatory aspects in clinical trials

Good Clinical Practice and Statistical Principles

- Declaration of Helsinki 1964: The World Medical Association (WMA) adopted a formal code of ethics for physicians engaged in clinical research
- International Council for Harmonization (ICH) guidelines follows the declaration of Helsinki
- *E6* Guideline for good clinical practice (GCP)
  - GCP is an international, ethical, scientific and quality standard for the conduct of trials that involve human participants
- *E9* Statistical Principles for Clinical Trials
- The aim of the ICH-GCP **Statistical Principles** is to demonstrate **Trustworthiness** in the results of clinical trials

Phases and evaluation structure of clinical trials

Phases	Pharmaceuticals	Is it randomized?
Phase I	**Safety**: Initial testing on human subjects	No
Phase II	**Prof-of-concept**: Estimating efficacy and optimal use on selected subjects	Sometimes
Phase III	**Comparison** against existing treatments in clinical settings	Yes
Phase IV	**Post-marketing** surveillance: For long-term side-effects	No

Statistical Principles and Standard Operation Procedures

The statistical principles and processes are describe in SOPs:

- **Statistical Study Planning**
  - Statistical design
  - Definition of primary and secondary endpoints
  - Statistical hypothesis and methods
  - Sample size determination
  - Randomization
  - Study populations
- **Randomization**
  - Specification of randomization process (single-blinded, double-blinded, open, stratification) in the study protocol
  - Generation of randomization lists
  - Implementation of the randomization (who?, where?, when?)
  - Unblinding process and serious adverse events (SAEs)

Statistical Principles and Standard Operation Procedures

- **Interim Analysis**
  - Specification of the interim analysis (design, time points, etc.)
  - Type of analysis (Efficacy, futility, sample size recalculation)
  - Methods (Group sequential, adaptive design, classical/Bayesian)
- **Statistical analysis plan (SAP)**
  - Detailed description of the statistical methods (e.g. missing data)
  - Relief of the study protocol from the detailed statistical methods
  - Connections between data management and statistics
  - Description of tables and figures

Statistical Principles and Standard Operation Procedures

- **Statistical report**
  - Detailed description of the processes that generate a statistical report
  - Template statistical report
  - Review processes and approval of the report
  - Data sources (closed study database, SAE reports)
- **Further working instructions**
  - Description and templates in R or other statistical sofware
  - Validation of the R code
  - Installation of software, etc.

Part 1:

Questions before the trial starts

Randomized Control Clinical Trial

Definitions

- A **clinical trial** is an experimental study comparing two (or more) medical treatments on human subjects most often patients
- When a **control group** is involved, this trial is called **controlled**
- For a **parallel** group design, one group of patients receives one treatment and the other group(s) receive(s) the other treatment
- All groups are **followed up** in time to measure the effect of the treatments

Randomized Control Clinical Trial

Methods against bias

- In a **randomized study**, patients are assigned to the treatments randomly
- To minimize **bias in evaluating the effect** of the treatments, patients and/or care givers are blinded
- When only patients are blinded one speaks of a **single-blinded** trial, but when both patients and care givers are blinded one speaks of a **double-blinded** trial

RCTs and the levels of clinical evidence

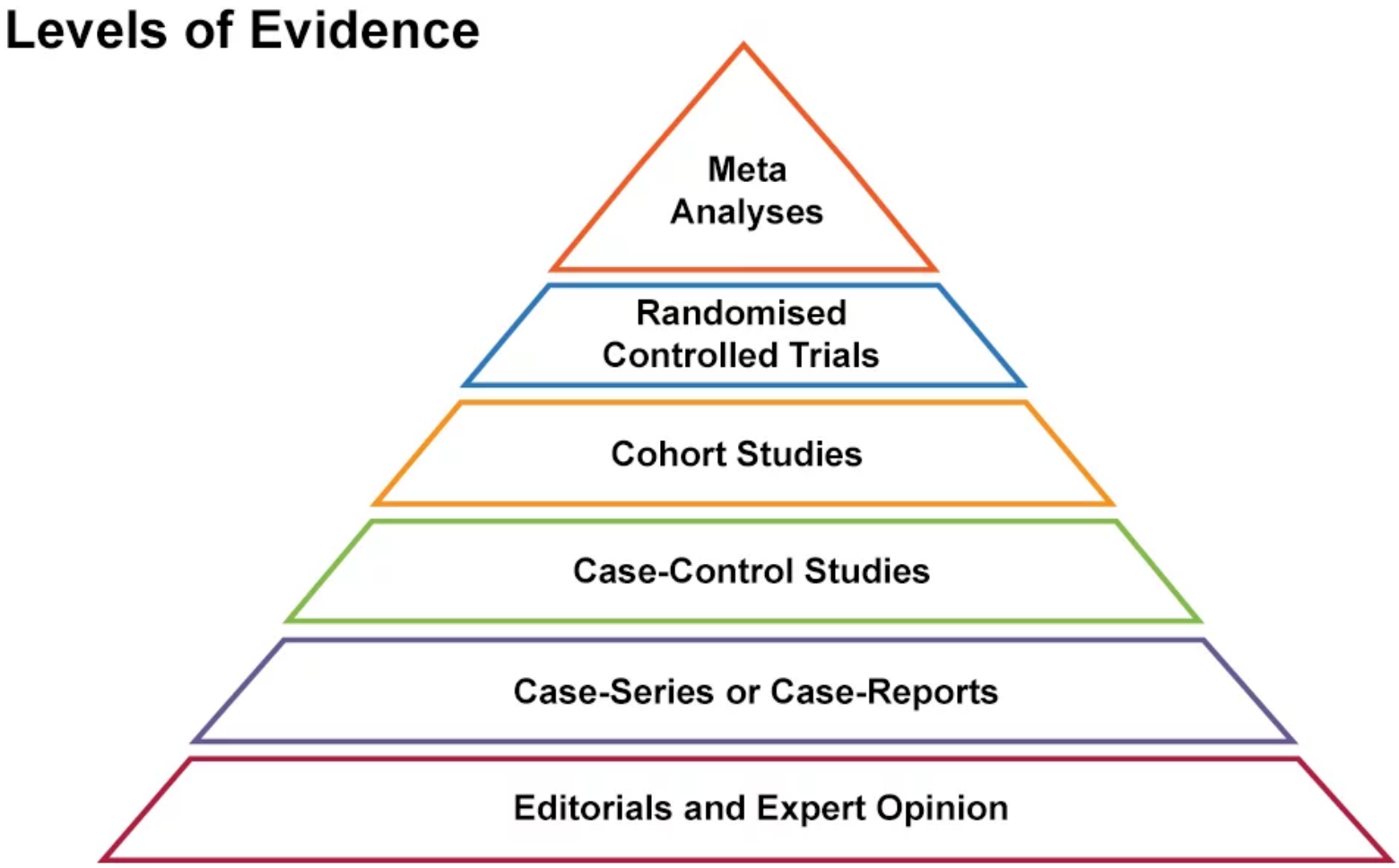


Figure 1: Levels of clinical evidence

Advantages of RCTs

- RCTs are the **gold standard** of clinical **evidence for efficacy**
- An RCT provides **unbiased estimates** of the treatment effect in the **sample at hand**
- RCTs have strong **internal validity**
  - RCTs avoid patient selection bias, i.e. the treatmeant assigment will not be based on patients' prognostic factors
  - RCTs control by confounders, i.e. randomization prevents effects from unknown prognostic factors
  - RCTs avoid errors in measurement of outcomes or information bias

Disadvantages of RCTs

- RCTs are costly and may take several years to accomplish
- RCTs suffers from **external validity**, i.e. the context of experimentation may not be the same as the context of clinical application
- The inclusion/exclusion criteria limit the **generalizability** of results
- A lack of **generalizability** is an issue if an RCT is used to make decisions in **real-world** medical practice

What constitutes a qualified statistician?

The ICH E9 Guideline on Statistical Principles states:

The responsibility for all statistical work associated with clinical trials will lie with an appropriately qualified and experienced statistician

(Gerlinger, C. et al. (2012) and Gelfonda J. A. L. et al.(2011))

- The **minimal education** needed to become a **qualified statistician** should be equivalent to at least a master's degree in statistics
- At least 3 years under the supervision of a senior statistician, before taking responsibility of a project
- The practical knowledge should include understanding of statistical methods, computational statistics, and statistical software

Sample size determination

ICH E9 Design considerations: Sample size

The number of subjects in the trial should always be large enough to provide a reliable answer to the questions addressed.

- This number is determined by the primary objective of the trial.
- Three components are used to determine this number:
  - A primary outcome variable
  - The null hypothesis and the clinical relevant effect
  - The significance statistical level of a test statistics

Example of sample size determination

- Primary outcome variable is a binary variable Y (success = 1 and failure=0)
- Parameter of interest is probability of success $Pr(Y = 1)$
- Two parallel treatment groups:
  - Control group: “standard medical care”
  - Treatment group: “new intervention”
  - Randomization: 1:1 (treatment:control)
- Clinical relevant effect:
  - Event rate of the control group 70%
  - Event rate of the treatment 85%

Example of sample size determination

- Clinical relevant effect:

```
p.control = 0.70
p.treatment = 0.85
# Difference
difference = p.treatment - p.control
difference*100
```

```
## [1] 15
```

```
# Odds ratio:
OR = (p.treatment/ (1-p.treatment)) /
      (p.control/(1-p.control))
OR
```

```
## [1] 2.43
```

- A difference of 15 % or an Odds Ratio of 2.43

Note Aside: Statistical significant testing

- An statistical hypothesis is a mathematical statement about a population parameter, this is called: **The Null Hypothesis** or H_0.
- The null hypothesis is contrasted with **the Alternative Hypothesis** or H_1, which indicates a particular departure from H_0
- In our example: "standard clinical care" vs. "new clinical care" correspond to:

$$H_0 : P_{Control} = P_{Treament} \quad \text{vs.} \quad H_1 : P_{Control} \neq P_{Treament}$$

- **A statistical test** is a decision procedure between H_0 and H_1
- The **p-value** measures how unusual is the observed data-set, given that the null hypothesis H_0 is correct, i.e. $Pr(Data|H_0 \; is \; True)$

Note Aside: Statistical significant testing

- In deciding to accept or reject the null hypothesis H_0, we might be making an "error"

		Type of Decision	
		Accept H_0	Reject H_0
True hypothesis:	H_0	Correct	Type I error
	H_1	Type II error	Correct

- The probability of making a *Type I* error is called α (e.g. α = 5%)
- The probability of making a *Type II* error is called β (e.g. β = 20%)
- The power is $1 - \beta$, which is the probability of correct rejection of H_0
- From these two type of errors, the Type I is fixed or controlled

Resulting sample size for the study

- We specify α = 0.05 and power $1 - \beta$ = 8.0
- We choose the statistical test: Two Proportion Z-Test (Normal approximation)
- The resulting sample size is:

```
##
##      Two-sample comparison of proportions power calculation
##
##                n = 120
##               p1 = 0.7
##               p2 = 0.85
##        sig.level = 0.05
##            power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

Note Aside: Drop-outs in RCTs

- Drop-outs in a *Randomized Control Trials* are very common
- Drop-outs are participants that:
  1) They meet inclusion criteria
  2) They are randomized to a treatment
  3) But they decided to quit the trial!
- From statistical perspective, a drop-out generates a **missing data**
- Drop-outs must not be confused with **censor data**, when the primary outcome of the study is **time to event** (e.g. survival)!

Sample size adjusted for drop-outs

- We take α = 5% and *Power* = 80% and the SSD results in N participants
- Suppose that a fraction of R_D participants could drop the study
- **Question:** Which is the total number N_{Total} of participants we need in the study?

$$N = N_{Total} - R_D \times N_{Total}$$

- The adjusted sample size for drop-outs is

$$N_{Total} = \frac{N}{(1 - R_D)}$$

Sample size adjusted for drop-outs

- In our example, if $R_D = 0.1$ then the total size N_{Total} is

```
# Sample size for both groups
# alpha = 5% and power = 80%

N
```

```
## [1] 241
```

```
# Drop out rate 10%:

R = 0.1

# Total sample size:

N/(1-R)
```

```
## [1] 268
```

Further remarks on sample size determination

- The SSD is **very sensitive** to the effect size
  - Example: We change the event rates to 71% (control) and 85% (treatment)

```
p.control = 0.71
p.treatment = 0.85

ssd2 = power.prop.test(p1 = p.control, p2 = p.treatment,
                       sig.level = 0.05, power = 0.8,
                       alternative = "two.sided")
N2 = ssd2$n
2*N2
```

```
## [1] 273
```

- In this example a 1% increase in the control group event rate increases the sample size by 32

Further remarks on sample size determination

- The SSD is **very sensitive** to the randomization plan
  - Example: Randomization 1:2

```
library(MESS)
N3 = sum(power_prop_test(p1 = p.control,
                 p2 = p.treatment,
                 ratio=2,
                 sig.level = 0.05,
                 power = .80)$n)

N3
```

```
## [1] 296
```

- If we randomize 1:2 the sample size increases by 24

Further remarks on sample size determination

- **Different statistical designs require different SSD:**
  - Cross-over trials (intra-individual correlation)
  - Cluster randomization (intra-cluster correlation)
- **Different outcomes require different SSD methods:**
  - Continues outcomes (e.g. mean differences)
  - Time to event outcomes (e.g. survival)
  - Ordinal cathegorical data (e.g. pain scores)
- **Some recommendations:**
  - Support the SSD with similar published studies
  - Whenever is possible support the SSD with a meta-analysis.

Randomization and imbalance assigments

- A **simple randomization** is a sequence of “flipping a coin” with 50/50 chances to assign *control* or *treatment*
- In general, **simple randomization** results in an imbalance assignment
- For example, $N = 100$, the chance of randomization yield exactly 50 subjects per group is only 7.96%
- Moreover, this imbalance also has impact in the distribution of the baseline characteristics of the subjects

Blocking randomization improves balance

- A block contains a pre-specified number and proportion of treatments assignments
- The size of the block must be an exact integer multiple of the number of treatments groups
- In practice, all blocks do not have the same size
- Varying the length of each block makes can help prevent discovery
- A sequence of blocks makes up **the randomization list**!

Effect of blocking randomization in real trials

	Amisulpride plus olanzapine group (n=110)	Amisulpride plus placebo group (n=109)	Olanzapine plus placebo group (n=102)	Total (n=321)
Age, years*	39·8 (12·1)	39·2 (10·8)	41·8 (12·3)	40·2 (11·7)
Sex†				
Male	77 (70%)	79 (72%)	73 (72%)	229 (71%)
Female	33 (30%)	30 (28%)	29 (28%)	92 (29%)
Ethnicity†				
White	100 (91%)	102 (94%)	94 (92%)	296 (92%)
Other	10 (9%)	7 (6%)	8 (8%)	25 (8%)
Weight, kg	85·2 (19·1)	79·1 (15·7)	80·9 (16·9)	81·7 (17·4)
Height, cm	175·1 (7·9)	175·4 (9·4)	175·4 (8·6)	175·3 (8·6)
Body-mass index, kg/m²	27·5 ± 6·1	25·7 ± 4·6	26·3 ± 5·2	26·5 ± 5·4
Partnered†	24 (22%)	18 (17%)	23 (23%)	65 (20%)
Employed†	27 (25%)	32 (29%)	25 (25%)	84 (26%)
Smoking				
Smokers†	91 (83%)	92 (84%)	68 (67%)	251 (78%)
Number of cigarettes per day‡	18·9 (11·8)	19·9 (10·1)	18·3 (8·8)	19·1 (10·4)
Diagnosis§				
Schizophrenia	95 (86)	94 (86%)	82 (80%)	271 (84%)
Schizoaffective disorder	14 (13%)	14 (13%)	19 (19%)	47 (15%)
Missing	1 (1%)	1 (1%)	1 (1%)	3 (1%)
Age at first psychiatric treatment, years*	26·3 (10·3)	25·2 (10·4)	27·9 (10·6)	28·2 (10·5)

Figure 2: COMBINE study with patients with schizophrenia. Baseline characteristics by treatments.

Effect of blocking randomization in real trials

Table 1: *Baseline Characteristics of study participants by Group. Data are presented as mean (SD) for continous variables and frequency (%) for categorical variables. P-values were obtained through independent samples t-test for continuous variables and chi-square test for categorical variables*

	level	off	on	p	test
n		25	23		
age (mean (SD))		47.92 (11.27)	50.82 (7.63)	0.314	
SEX (%)	female	11 (44.0)	7 (30.4)	0.502	
	male	14 (56.0)	16 (69.6)		
country (%)	austria	4 (16.0)	4 (17.4)	1.000	exact
	france	3 (12.0)	2 (8.7)		
	germany	17 (68.0)	16 (69.6)		
	swizerland	1 (4.0)	1 (4.3)		
HEIGHT (mean (SD))		173.80 (9.24)	174.43 (7.50)	0.796	
WEIGHT (mean (SD))		66.58 (13.39)	69.38 (10.53)	0.427	
ETHNIC (%)	caucasian	25 (100.0)	23 (100.0)	1.000	exact
	non-caucasian	0 (0.0)	0 (0.0)		
HAND (%)	left	2 (8.0)	0 (0.0)	0.490	exact
	right	23 (92.0)	23 (100.0)		

Figure 3: HDDBS patients with Huntington Disease: Baseline characteristics by treatment. Group off is control and group on is Deep Brain Stimulation.

The Study Population and the Intention to Treat Principle

- Golden principle in statistics: **The statistical analysis must be perform according to randomization**
- **The Intention to Treat (ITT)** analysis includes all randomized patients that received a treatment regardless of what treatment (if any) they received
- This method allows the investigator **to draw accurate (unbiased) conclusions** regarding the effectiveness of an intervention
- For discussion about **ITT** vs. **Per-protocol (PP)** analysis see: Dunn et al. Trials (2018) 19:499 https://doi.org/10.1186/s13063-018-2885-z

Part 2:

Questions during the trial

Monitoring and sequential trials

- **An early stop of a trial** is a quite complex ethical, financial, and scientific issue, in which statistical analysis plays an important role
- **Data-depending** stopping may have several reasons (Piantadosi, 1997, pag. 231) like:
  - Treatments are found to be convincing different (or not different) by experts.
  - Side effects or toxicity is too severe.
  - The data are of poor quality.
  - Accrual of patients is too slow to complete the study.
  - The scientific questions are no longer important because of other developments.
  - etc . . .

Monitoring and sequential trials

- Recommendations concerning early stopping rest in the hands of independent committees known as **data monitoring committees (DMC)**
- From the statistical point of view there are several techniques to design and to monitor a trial.
- The most commonly used are:
  - **Group sequential designs**
  - **Adaptive designs**
  - **Bayesian statistical techniques**

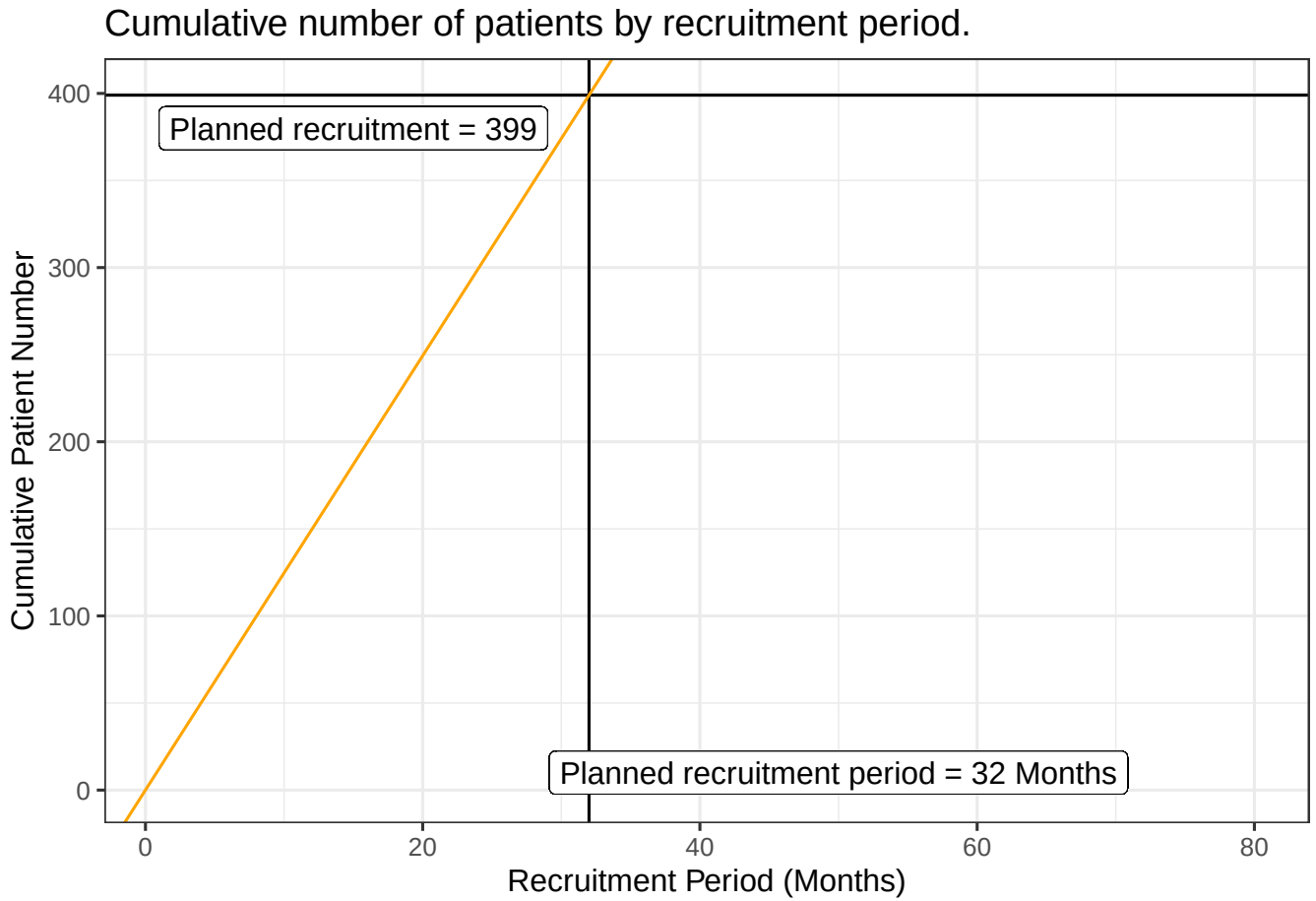
Monitoring and sequential trials

- **Group sequential designs**
  - The most popular approach for design and monitor a trial
  - Usually, very simple to implement
- **Adaptive designs**
  - They allow more than just stopping at an interim analysis
  - They can be considered as flexible way to combine multiple analysis at interim
- **Bayesian statistical techniques**
  - Conceptually clear by the sequential used of the Bayes theorem. For more information see Spiegelhalter et al. 2004

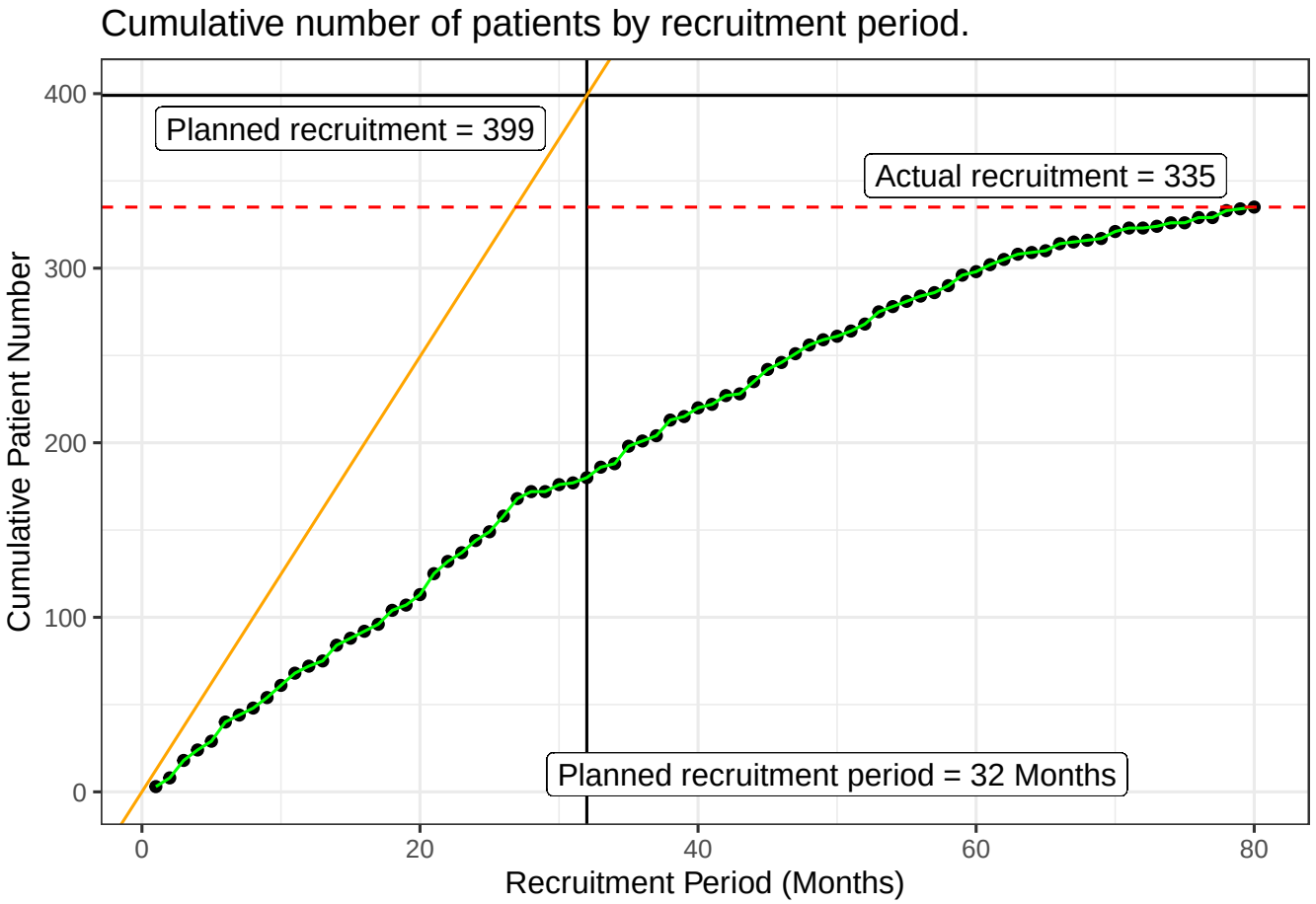
Patient recruitment

- **The Lasagna Law:** The number of patients available to join a trial drop by 90% the day the trial begins. They reappear as soon as the study is over
- Delayed patient recruitment is such a **well-known problem** that it has been used to **detect fake RCTs** publications (Gaby, A. R. (2022) Integrative Medicine Vol 21, No. 2)

Patients recruitment: multi-center RCT

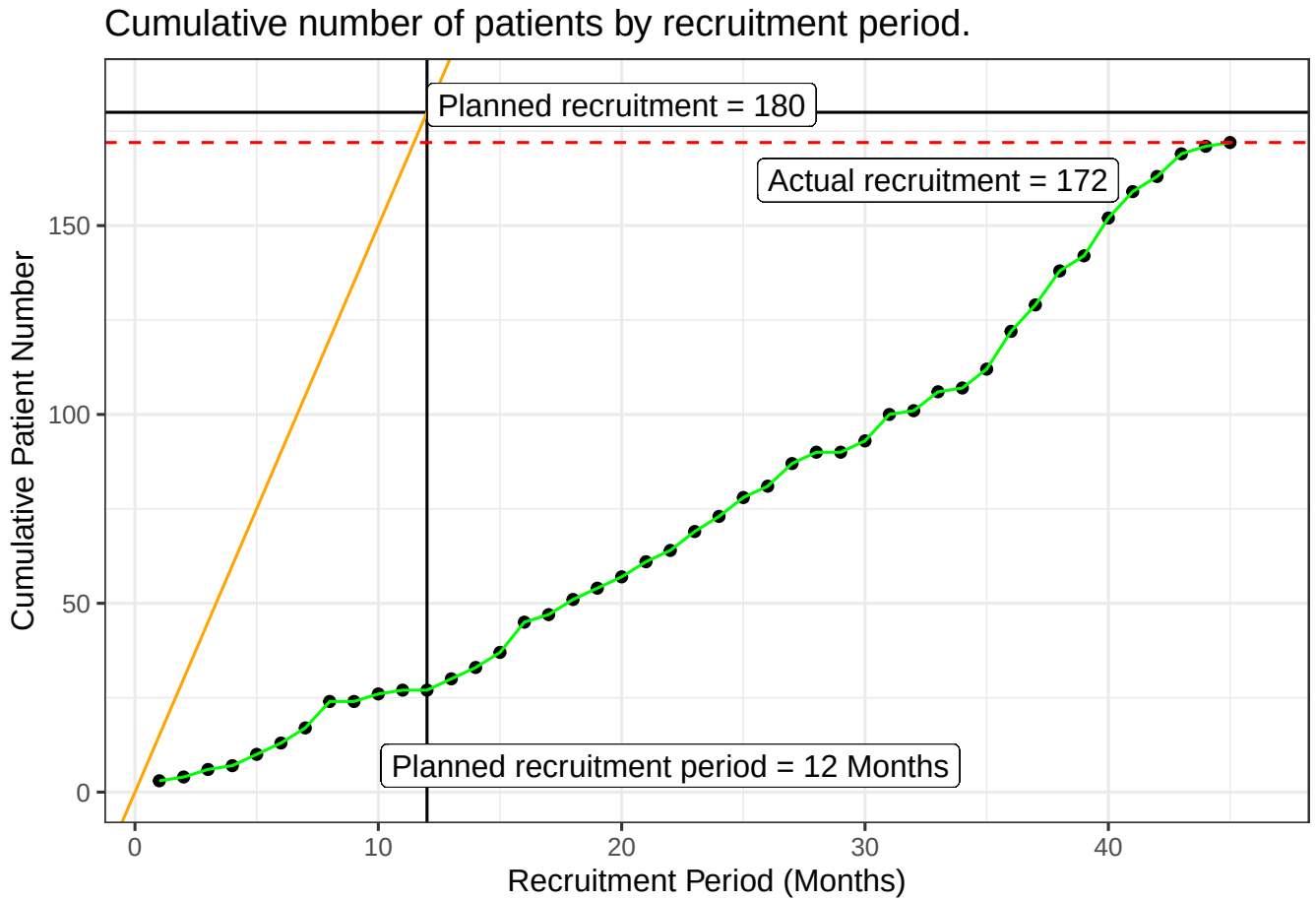


Patients recruitment: multi-center RCT



Patient recruitments: multi-national and center RCT

Patient recruitments: multi-national and center RCT



Part 3:

Questions after the trial data

Statistical analysis and reporting with R

What is R?

- R is a free and open source software for any kind of analysis involving data
- R is a dialect of **S** system!

What is S?

- The **S System** was a 20 years software research project at the Bell Labs in the USA led by **John Chambers**
- In 1998 **John Chambers** was awarded by the Association for Computing Machinery (the “Nobel Prize” in Computer Science)
- Previous awards have included cornerstones of modern computing such as the UNIX system, the World Wide Web, TCP/IP, and the Postscript language

A replicability enviroment with R

- **The replication crisis** is an ongoing methodological crisis in which it has been found that the results of many scientific studies are difficult or impossible to reproduce
- **Dynamic reporting software** in **R** have supported a replicability research environment
- The idea is to have both **computing code** and **narratives** in the same document
- Results are generated automatically from the source code

A replicability enviroment with R

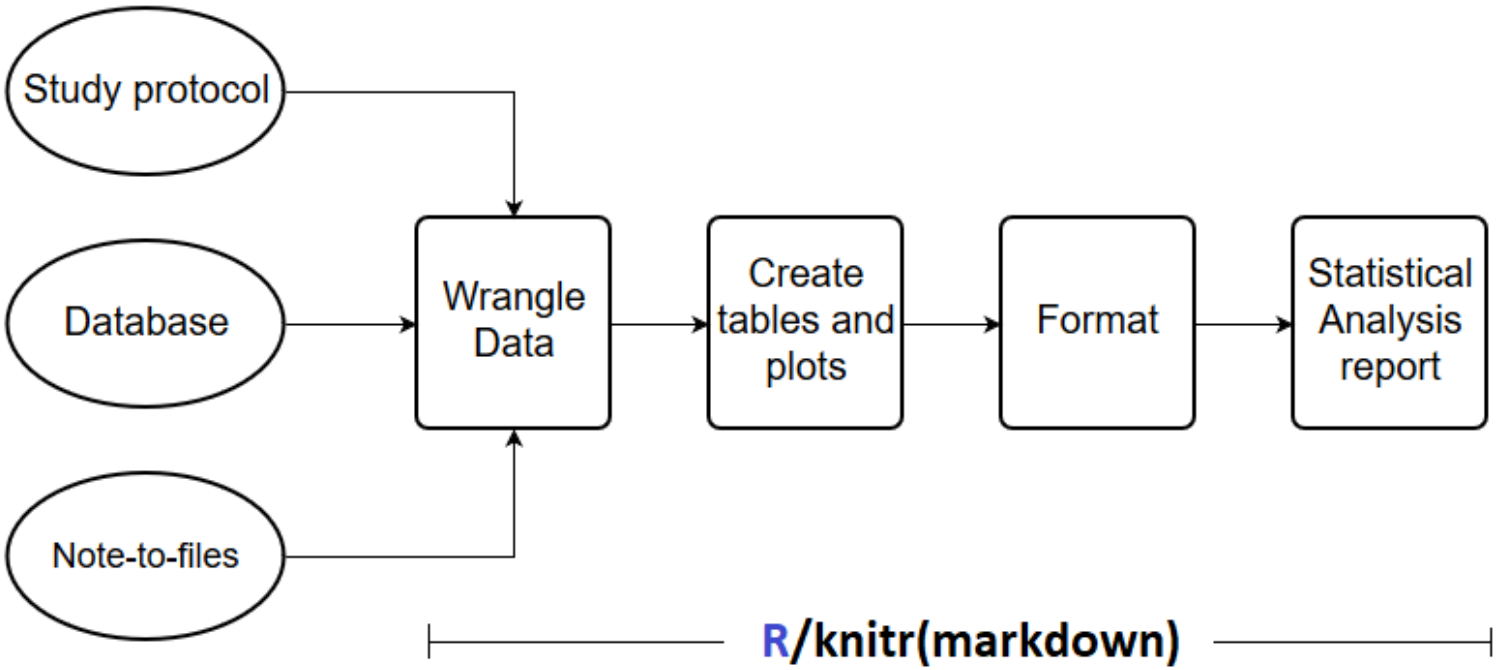


Figure 4: Our workflow for combining R scripts with Markdown

Issues with outliers

- What is an outlier?
  - Outliers are valid data
  - Outliers are not errors in RCTs' data-set
  - Outliers are extreme values that stand out from the overall pattern of values in a data-set
- What is the influence of **a single outlier**?
  - A single outlier can completely change results
  - A single outlier can bias results, increase variability or both

How to deal with with outliers

- It is impossible to account for this kind of problems at the design of the trial
- In general we need to have a preliminary export of the data
- We can use graphical and exploratory methods to detect outliers
- Robust (Bayesian or classical) statistical methods should be used to analyze data with outliers
- Robust methods reduce the influence of outliers
- Robust methods are efficient, i.e. without outliers give the similar result as classical ones

Real data analysis without outliers

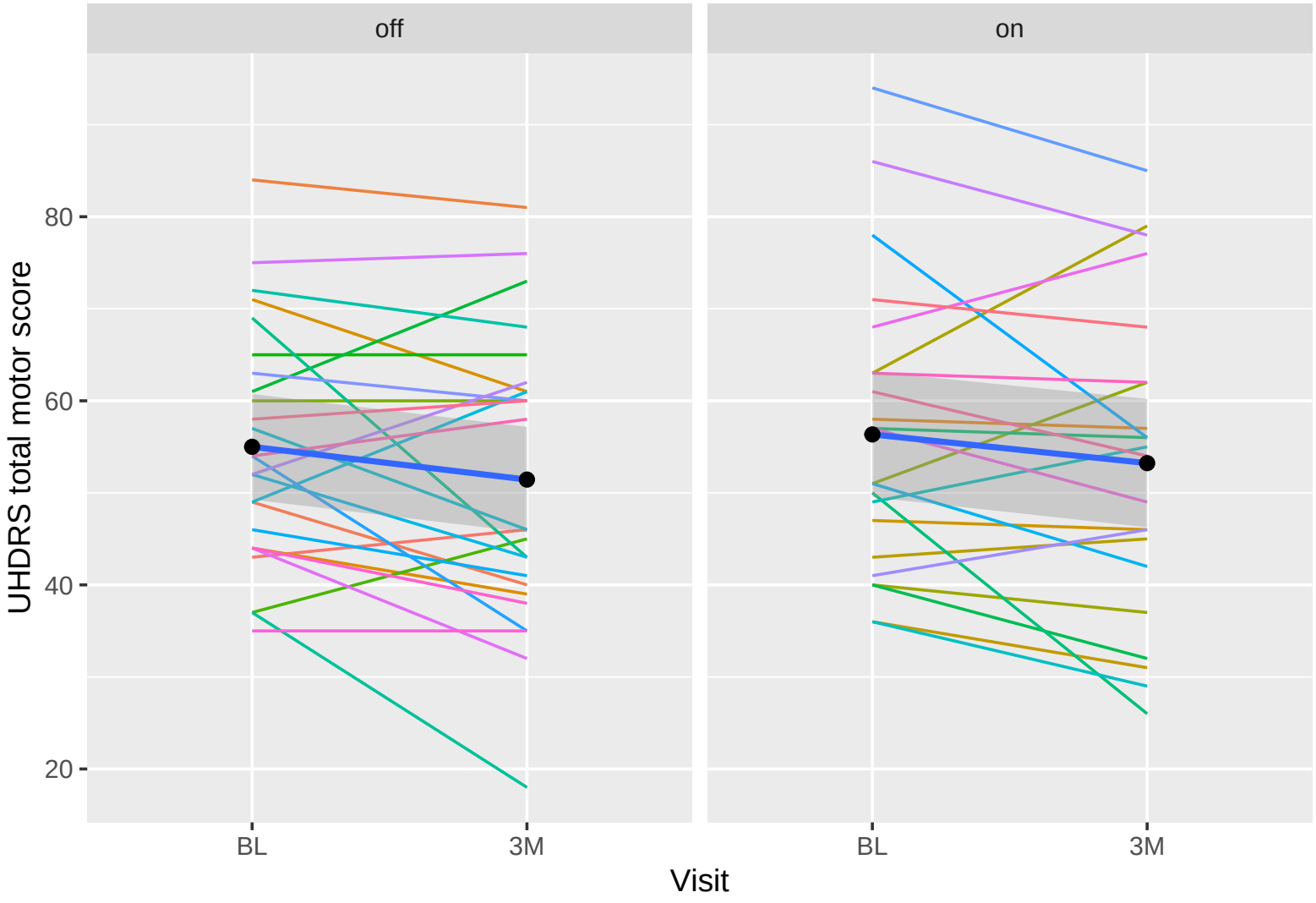


Figure 5: Patients profiles from baseline to three months.

Real data analysis without outliers

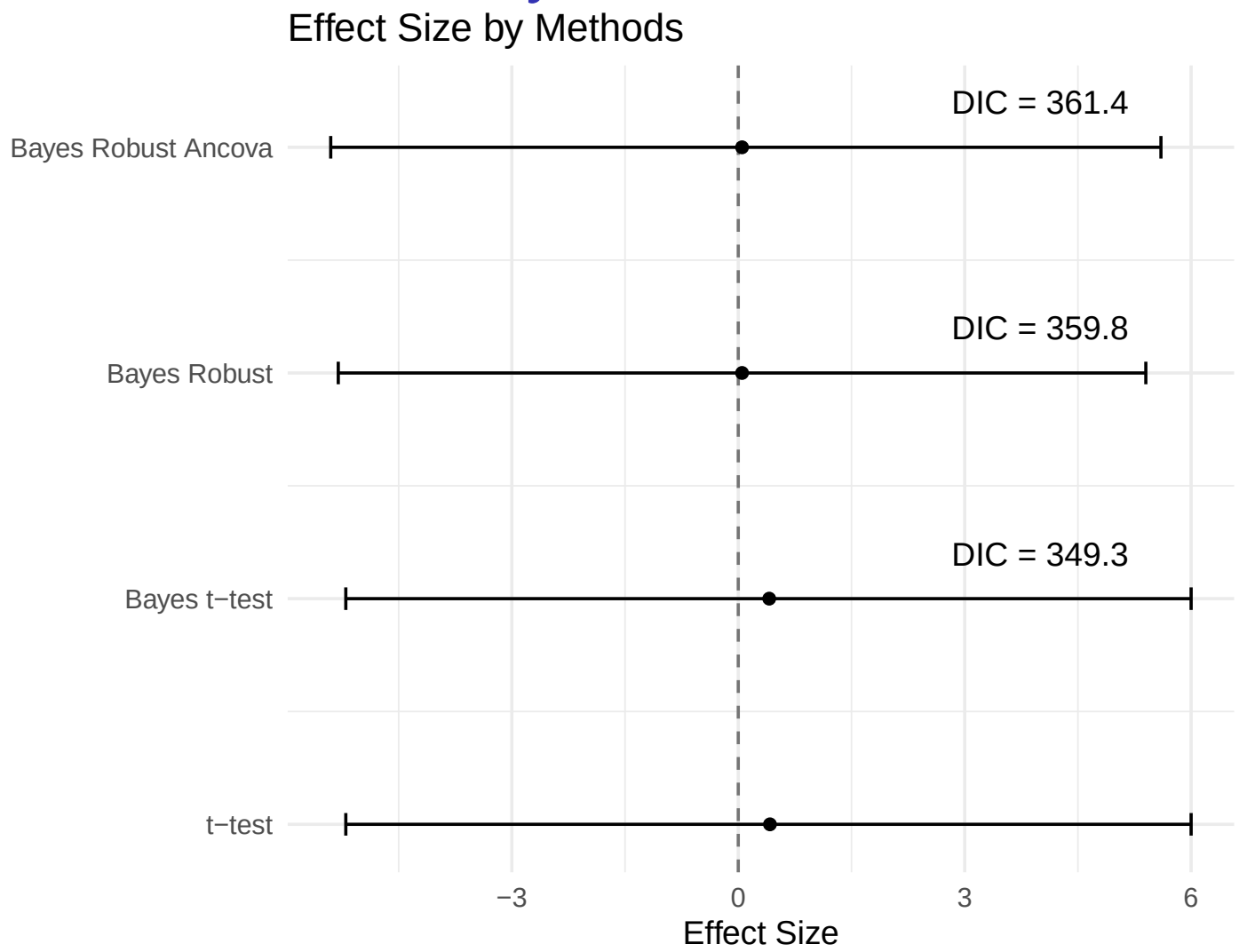


Figure 6: Statistical analysis: The Bayesian Robust methods give similar results compare to a t-test (DIC the lowest the better).

Real data analysis with outliers

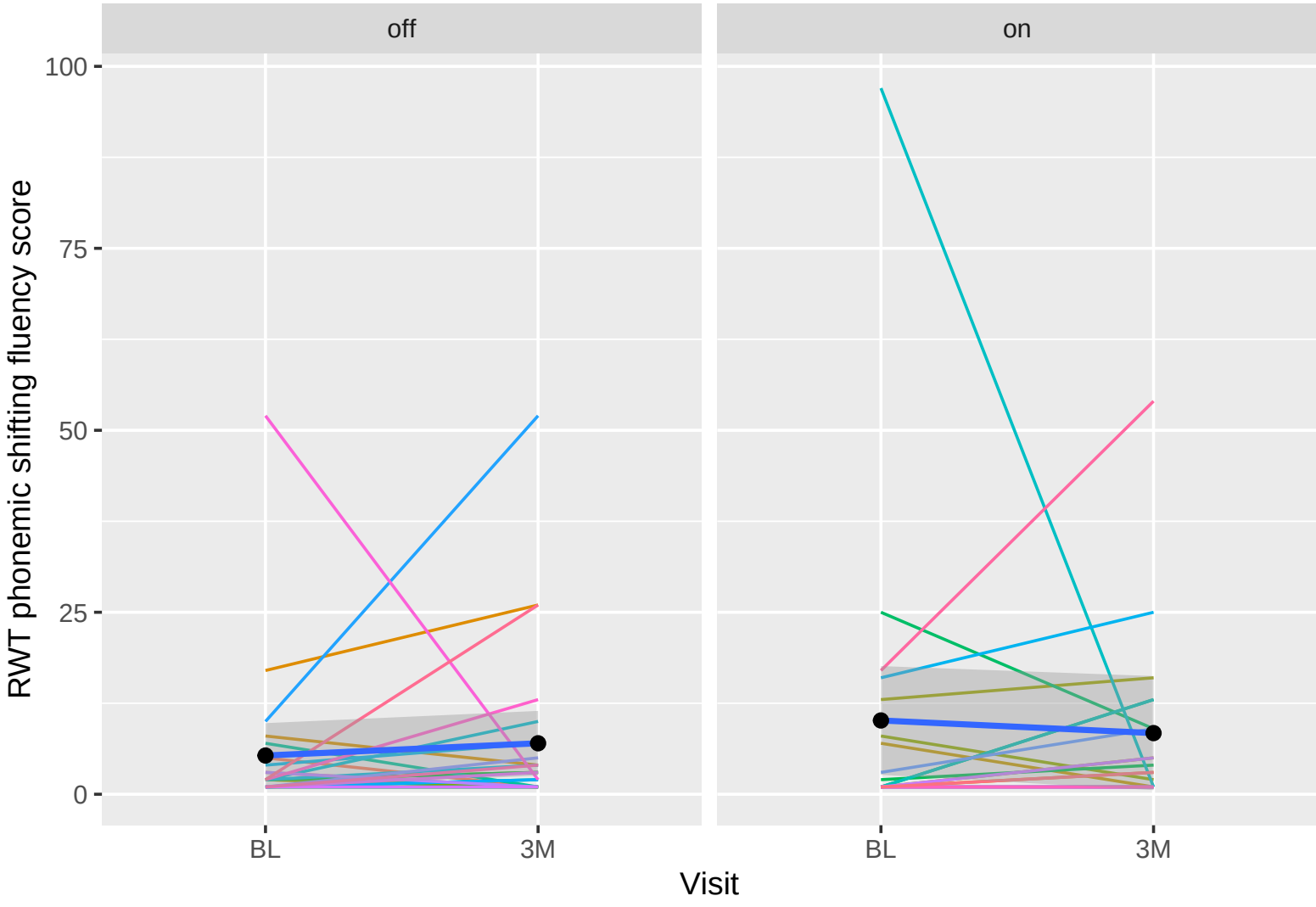


Figure 7: Patients profiles from baseline to three months.

Real data analysis with outliers

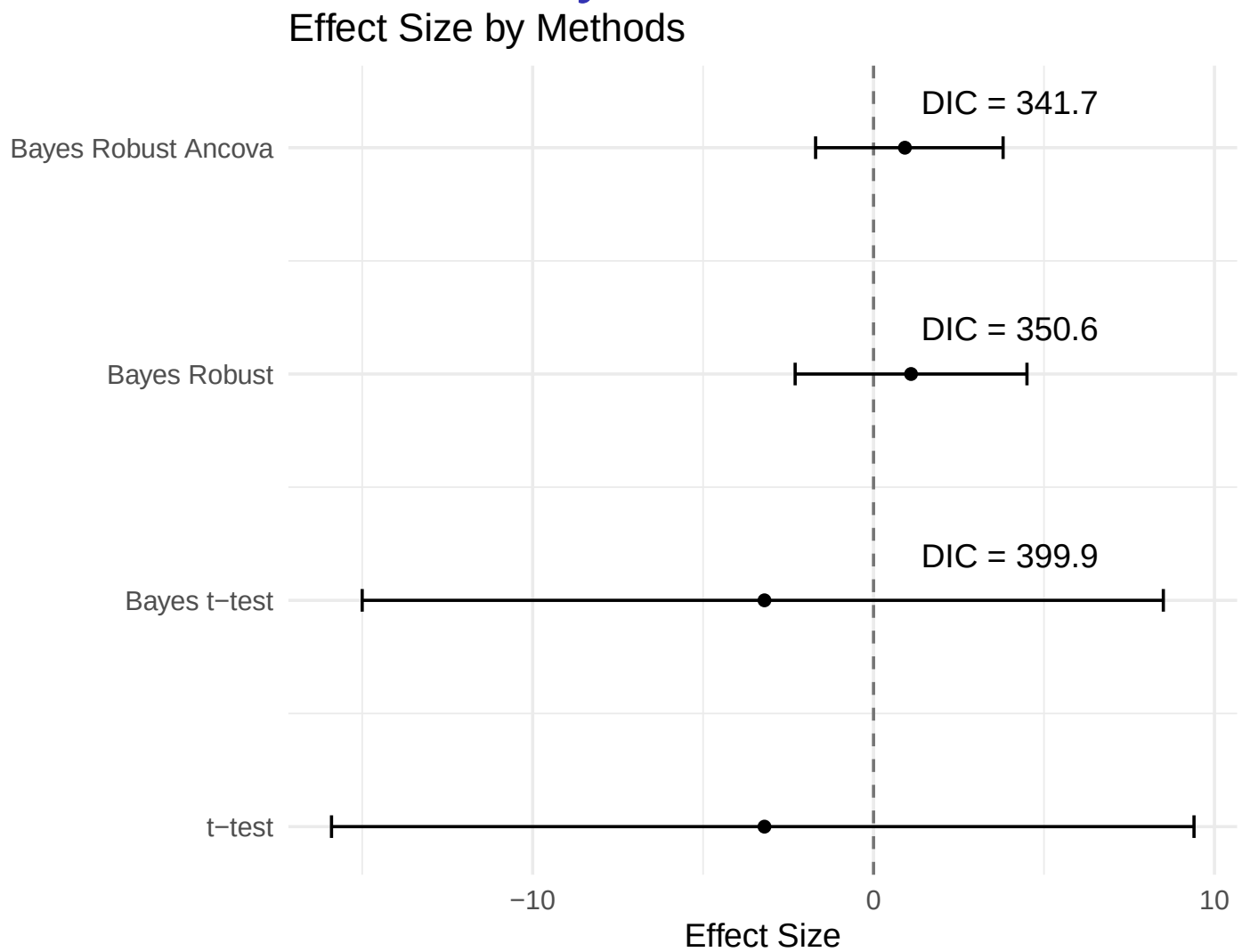


Figure 8: Statistical analysis: The Bayesian Robust methods give strong corrections compare to a t-test (DIC the lowest the better).

An example of longitudinal data with outliers

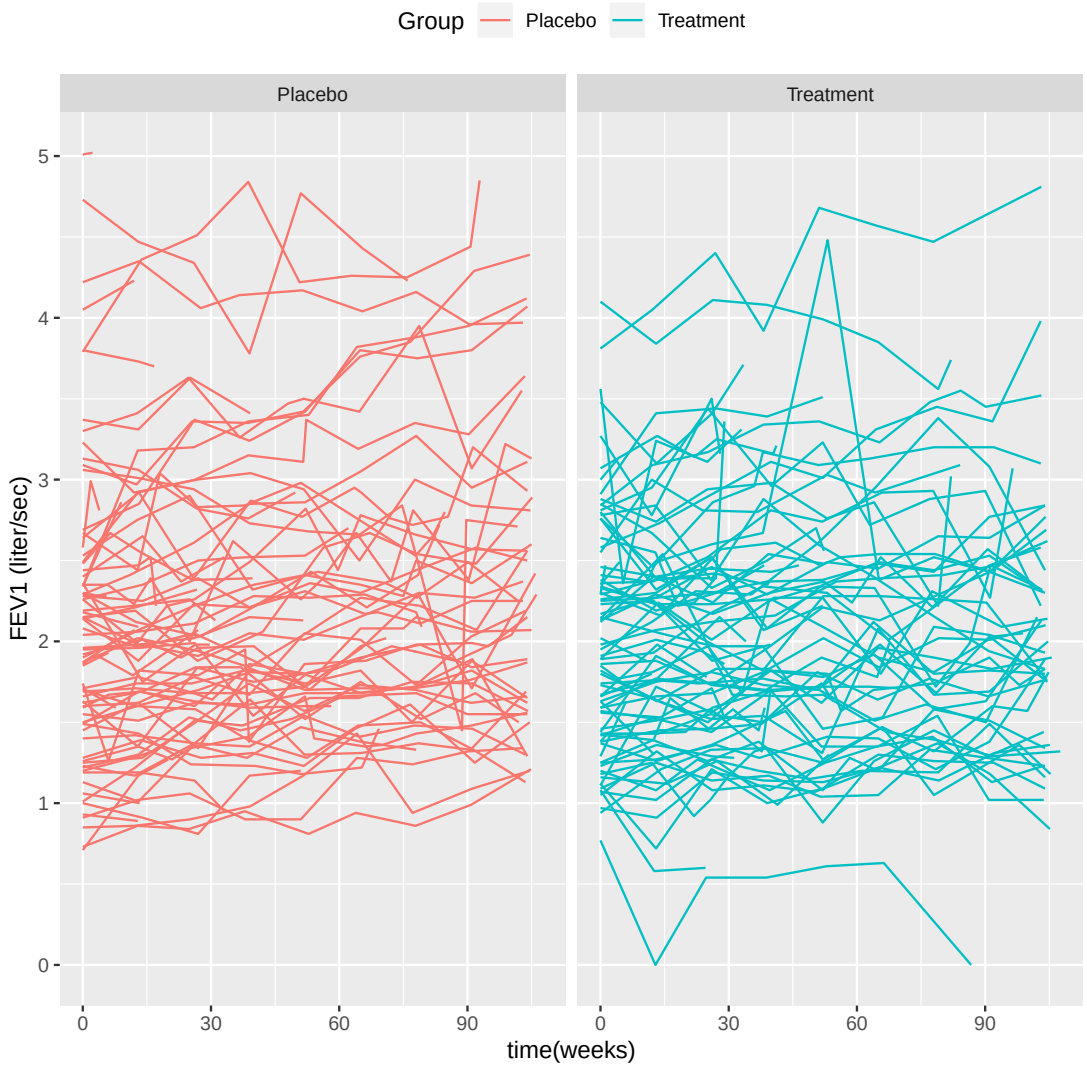


Figure 9: Forced expiratory volume in litters per second. Patients profiles in weeks.

Detection of outliers by fitting a mixed-effects model

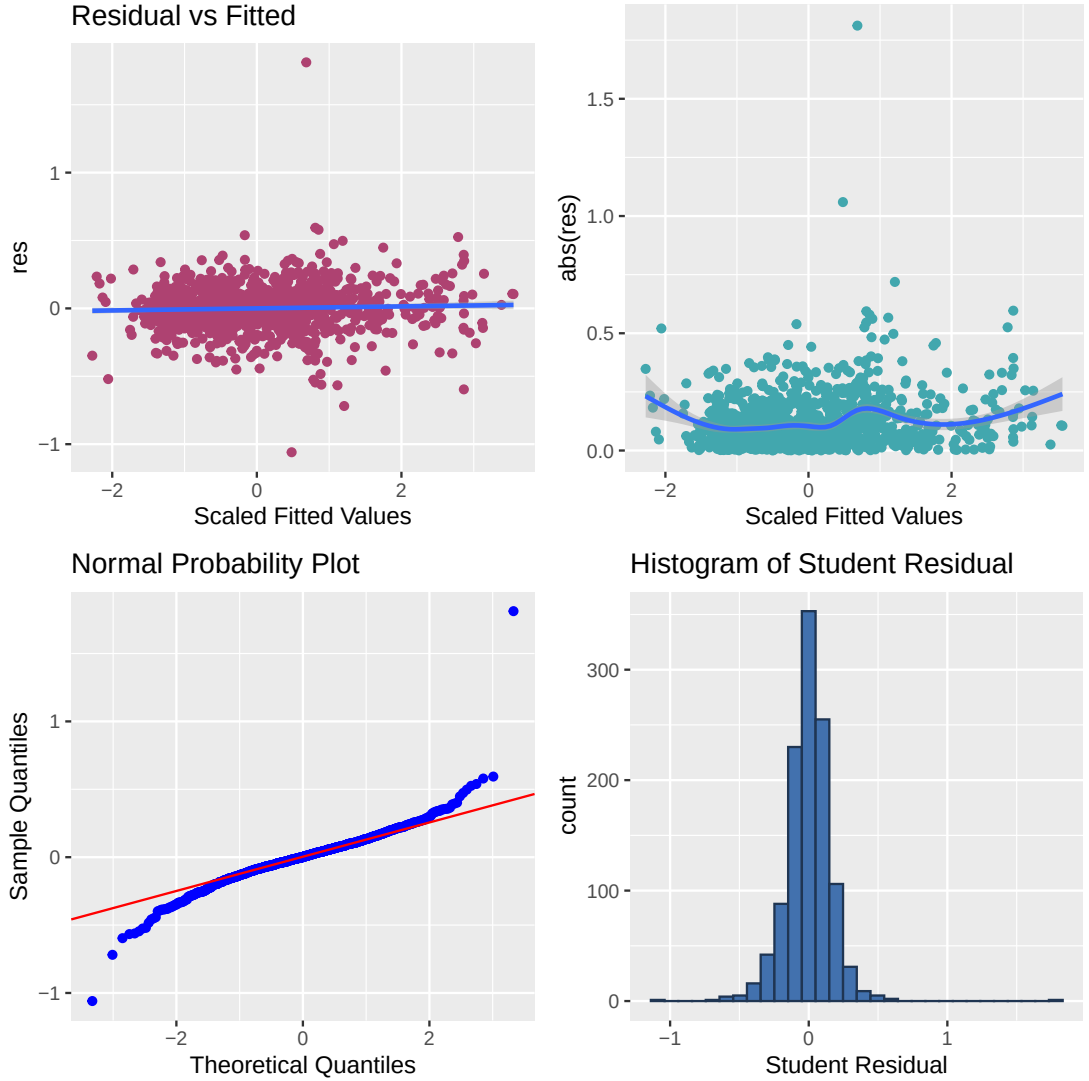


Figure 10: Model diagnostics and outlier detection after fitting a mixed effect model with Normal random-effects.

Robust Bayesian model for longitudinal data

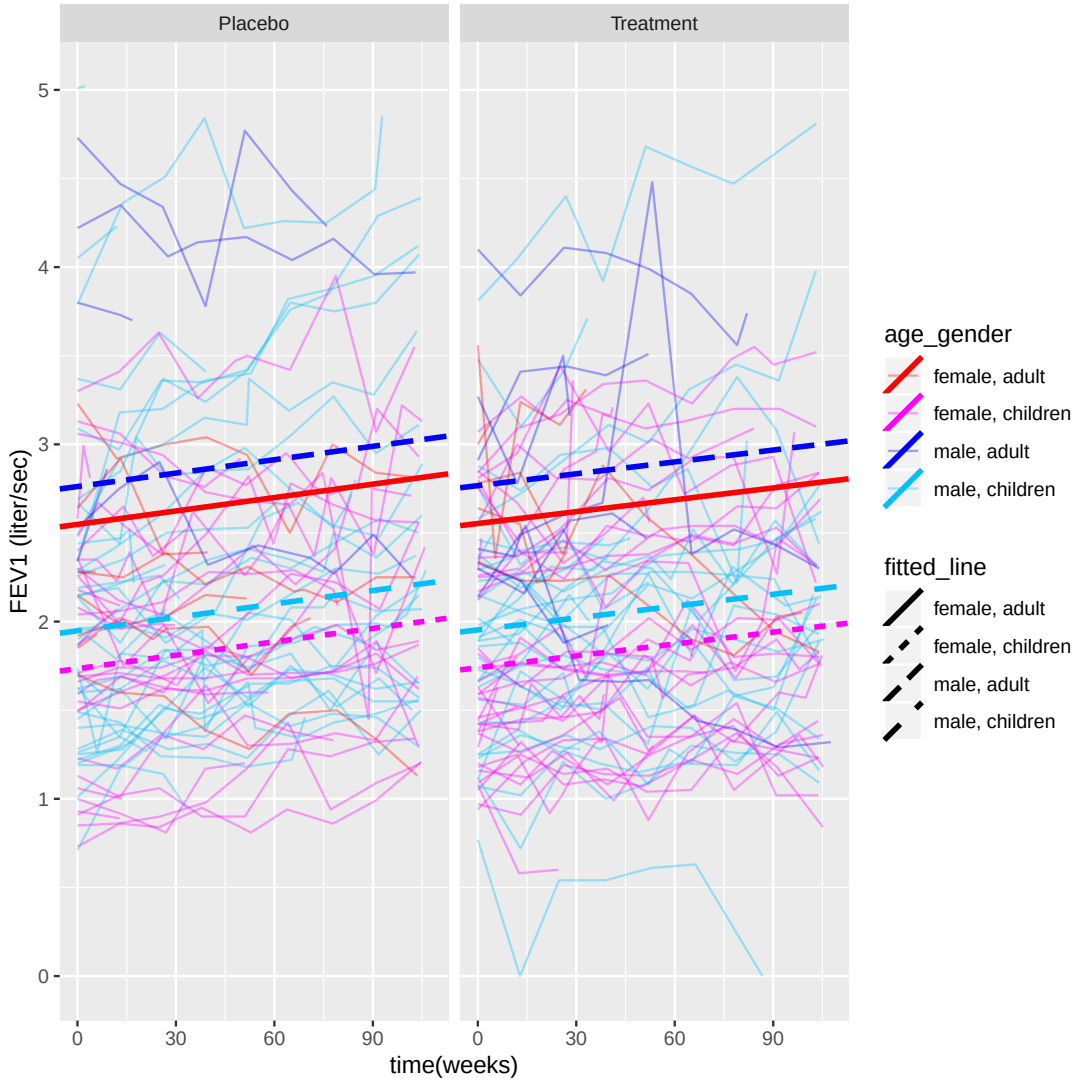


Figure 11: Results: Fixed-effects after fitting a Bayesian Robust model with Non-Normal random-effects.

Issues with missing data in RCTs

- Missing data are common in RCTs
- Missing data is a threat to the validity of treatment effect estimates in RCTs
- Missing data is usually inadequately handled in both RCTs and observational research
- **A large gap** exists between **statistical methods research** related to missing data and **use of these methods** in RCTs in top medical journals!

Types of missing data: MCAR, MAR, MNAR

The classification of different types of missing data is based on **the chance** of an observation being missing:

- **Missing Completely at Random (MCAR):** The chance of a missing observation does not depend on observed or **unobserved data**
- **Missing at Random (MAR):** The chance of a missing observation does NOT depend on the **unobserved value**, but may depend on other observed data
- **Missing Not at Random (MNAR)** The chance of a missing observation depends on the **unobserved value**
- These categories are a continuum: strength may vary

Dealing with missing data

- It is important to perform descriptive statistics and exploratory analysis to understand possible explanations for missing data
- Statistical inference is limited by the fact that it is **not possible to validate** if the missing data mechanisms is MAR or MNAR
- Bayesian sensitivity analysis centers the analysis at MAR and generates different scenarios in directions to MNAR
- If different scenarios do not change conclusions then we can claim treatment effects

Part 4:

Should we trust in medical algorithms?

Trustworthiness and algorithms

In **silico clinical trials** are medical algorithms the point is:

Should we trust in medical algorithms? (Spiegelhalter, 2020)

A generic check list from FATML (Fairness, Accountability and Transparency in Machine Learning)

- **Responsibility**: whom to approach when things go wrong
- **Explainability**: to stakeholders in nontechnical terms
- **Accuracy**: to identify sources of error and uncertainty
- **Auditability**: to allow third parties to check and criticize
- **Fairness**: to different demographics

Trustworthiness and medical algorithms

- A check list like a FATML assumes that **algorithms will be beneficial**
- In clinical applications, we have to consider **Impact**: What are the **benefits** and **harms** in actual use?
- The **impact** of a medical algorithm can be evaluated in a four-phase structure (Spiegelhalter, 1983 and Stead, et al. 1994)

Phases and evaluation structure of clinical algorithms

Phases	Pharmaceuticals	Algorithms
Phase I	**Safety**: Initial testing on human subjects	**Digital testing**: Performance on test cases
Phase II	**Prof-of-concept**: Estimating efficacy and optimal use on selected subjects	**Laboratory testing**: Comparison with humans, user testing
Phase III	**Comparison** against existing treatments in clinical settings	**Field testing**: Controlled trials of **impact**
Phase IV	**Post-marketing** surveillance: For long-term side-effects	**Routine use**: Monitoring for problems

Phases and evaluation structure of clinical algorithms

- Most of the attention in the published literature has focused on **Phase I**, which the claimed accuracy in on digital data sets (Spiegelhalter, 2020).
- There is an increase **Phase 2** evaluations in diagnostic tests where algorithms and human experts results are independently assessed by experts (Dick, et al. 2019)
- There are few **prospective Phase 3 validation** for tasks that algorithms could help clinicians that would be useful for health systems (Topol, 2019).

The RAPT meta-analysis

- The Risk Assessment and Prediction Tools (RAPT)
- Systematic reviews of clinical decision tools for acute abdominal pain (Liu, et al. 2006)
- Include 13 **prospective phase 3 diagnostic studies** comparing diagnostic accuracy of **Doctors** vs **Doctors with Added Tools**
- **Added Tools**: 4 Neural-Nets; 7 Bayesian; 2 Logistic regression
- The data and Bayesian meta-analysis available in **bamdit** (Verde 2018)

The RAPT meta-analysis

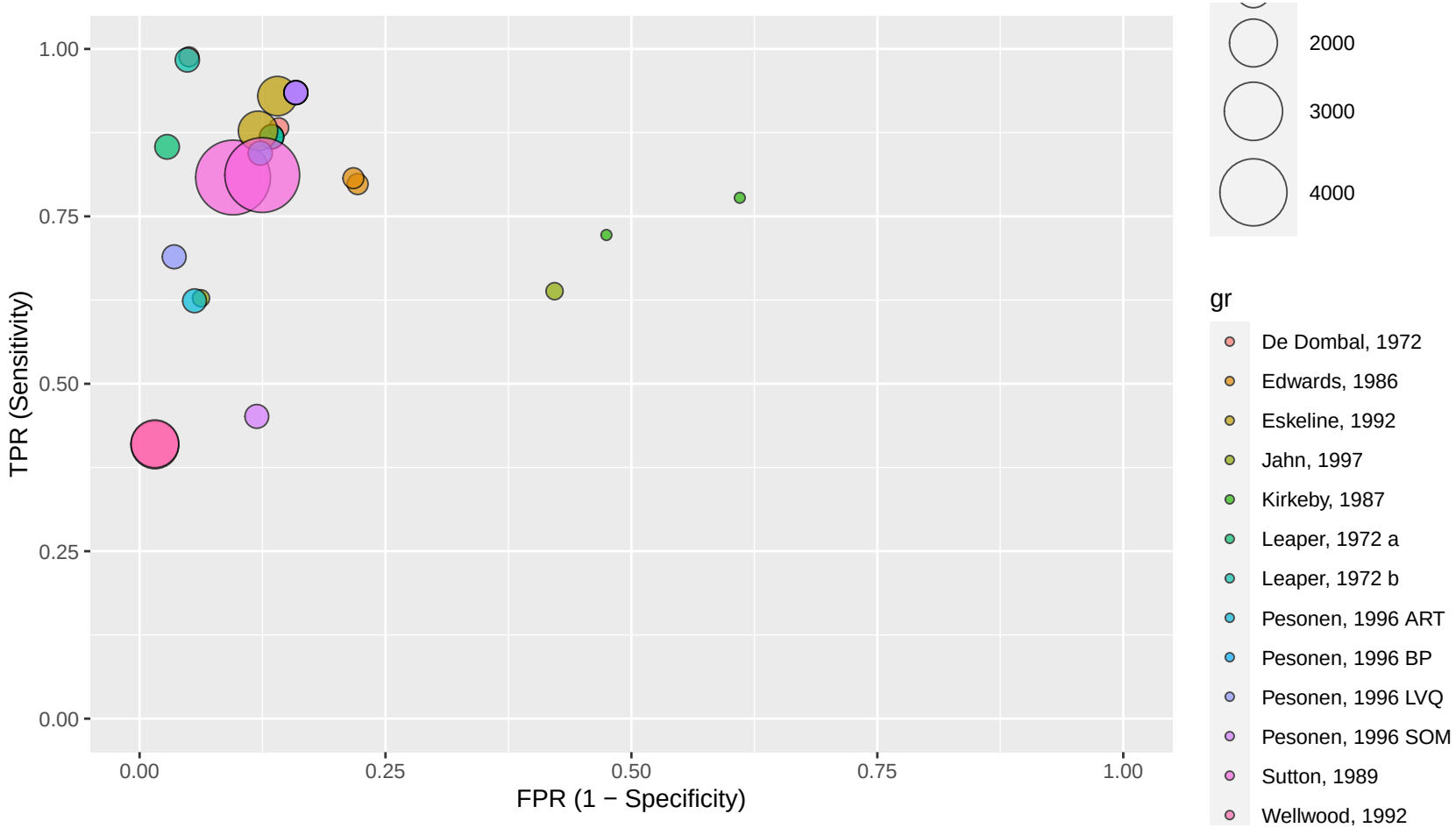


Figure 12: Meta-Analysis of accute abdominal pain comparing doctors vs doctors+tools

The RAPT meta-analysis

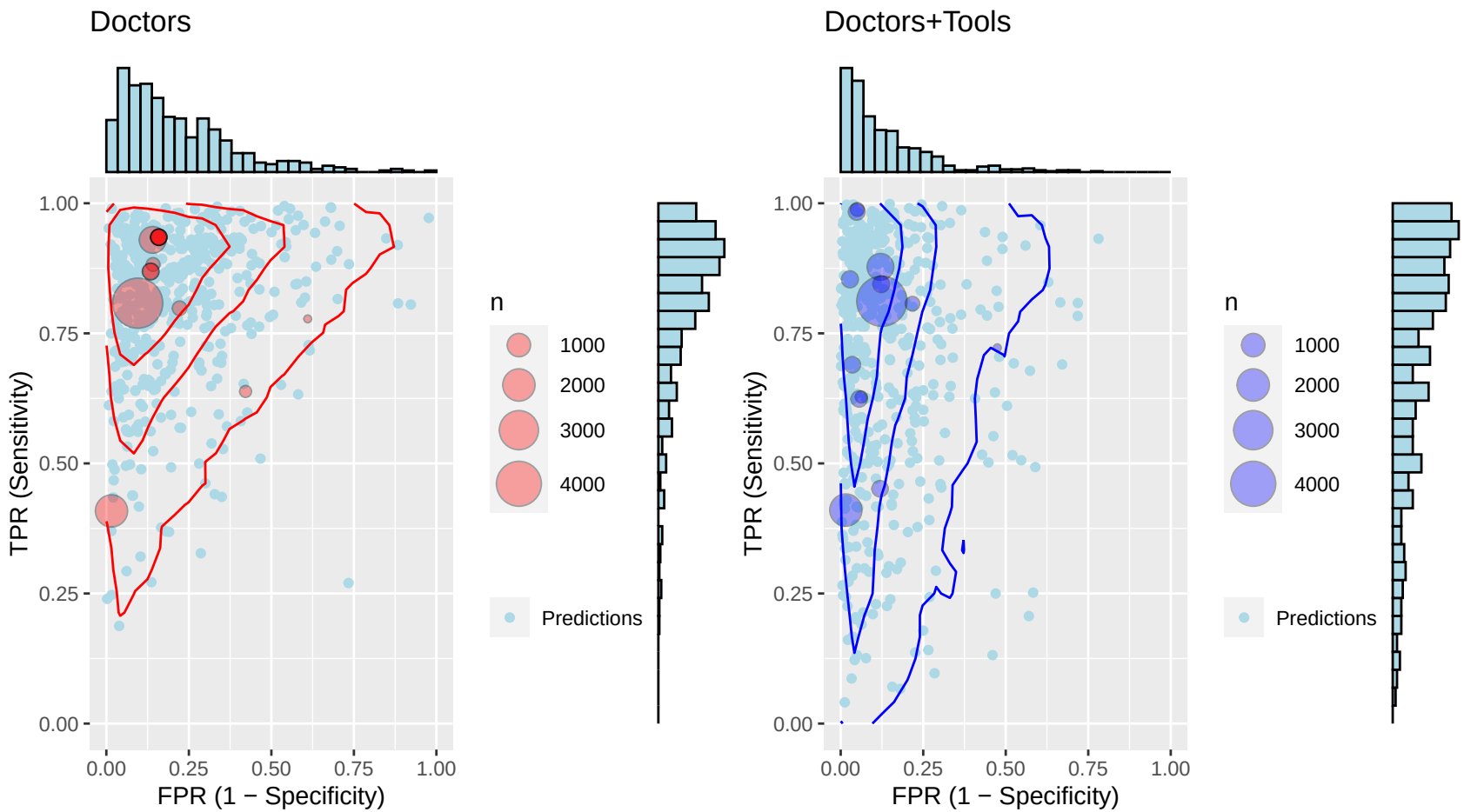


Figure 13: Meta-Analysis results: Left panel accuracy of doctors. Right panel: doctors+tools

The RAPT meta-analysis

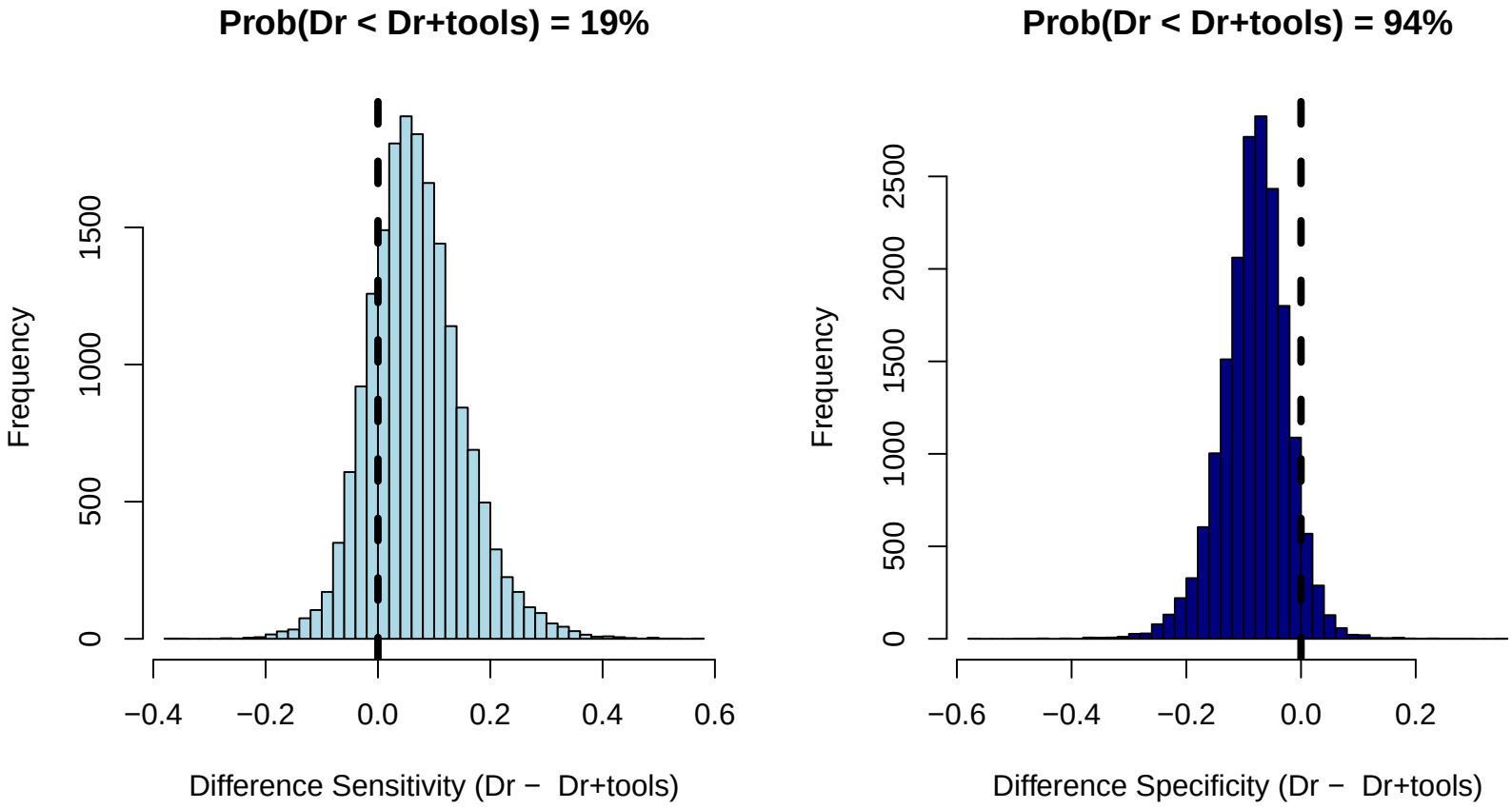


Figure 14: Meta-Analysis results. Left panel: difference in sensitivity. Right panel: Difference in specificity

Conclusions

- Trustworthiness in medical algorithms should increase by following a similar evaluation structure as other clinical interventions
- Evaluation of medical algorithms in phase 3 controlled trials allows to assess the benefit and harms in the real applications
- Assessment of uncertainty of an algorithm in clinical real-world environment also increase the its trustworthiness
- The infrastructure used in classical Randomized Control Trials applies to the evaluation of medical algorithms

Thank you very much!!

Vielen Dank!!

Muchas gracias!!

Acknowledgements

- For permission of using the data of RCTs from the Faculty of Medicine at HHU
  - Prof. Dr. med. Jan Vesper (HDDBS Study on huntington)
  - PD Dr. med. Christian Schmidt-Kraepelin (COMBINE Study on schizophrenia)
  - Prof. Dr. med. Antje Schuster (IMPACT Study on cystic fibrosis)

Bibliography

- Dunn, D.T., Copas, A.J. and Brocklehurst, P. Superiority and non-inferiority: two sides of the same coin?. Trials 19, 499 (2018). https://doi.org/10.1186/s13063-018-2885-z

- Friedman L.M., Furberg C.D. and DeMets, D.L. (1998) *Fundamental of Clinical Trials*. Springer Verlag, New York, 3th Edition.

- Gaby, A. R. (2022) Is There an Epidemic of Research Fraud in Natural Medicine. Integrative Medicine Vol 21, No. 2

- Gerlinger, C. et al. (2012) Considerations on what constitutes a qualified statistician in regulatory guidelines. Stat. Med., 31, 1303–1305

- Gelfonda J. A. L. et al.(2011) Principles for the ethical analysis of clinical and translational research. Stat. Med. 15; 30(23): 2785–2792. doi:10.1002/sim.4282.

- ICH E6 (R3) (Good Clinical Practice 2023)

- ICH E9 (Statistical Principles for Clinical Trials 1998)

- Liu JL1, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, Ohmann C, Wellwood J, Dawes M, Altman DG. Health Technol Assess. (2006) Nov;10(47):1-167, iii-iv. Systematic reviews of clinical decision tools for acute abdominal pain.

Bibliography

- Piantadosi S. (1997). *Clinical Trials: A Methodological Perspective*. Wiley Series in Probability and Statistics.

- Schmidt-Kraepelin, C. et al. Amisulpride and olanzapine combination treatment versus each monotherapy in acutely ill patients with schizophrenia in Germany (COMBINE): a double-blind randomised controlled trial, The Lancet Psychiatry, Volume 9, Issue 4, 2022, Pages 291-306, https://doi.org/10.1016/S2215-0366(22)00032-3.

- Spiegelhalter, D. (2020). Should We Trust Algorithms? Harvard Data Science Review, 2(1). https://doi.org/10.1162/99608f92.cb91a35a

- Spiegelhalter DJ, Abrams KR and Myles JP (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* Wiley, Statistics in Practice.

- Stead, W. W., Haynes, R. B., Fuller, S., Friedman, C. P., Travis, L. E., Beck, J. R.,Abola, E. E. (1994). Designing medical informatics research and library–resource projects to increase what is learned. Journal of the American Medical Informatics Association, 1(1), 28–33. https://doi.org/10.1136/jamia.1994.95236134

Bibliography

- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

- Verde, P. E. (2018). bamdit: An R Package for Bayesian Meta-Analysis of Diagnostic Test Data. Journal of Statistical Software, 86(10), 1-32. doi:10.18637/jss.v086.i10