Thinking beyond the norm: how to (fairly) evaluate Bayesian Dynamic Borrowing designs

Gaëlle Saint-Hilary (Saryga)
On behalf of Simon Wandel (Novartis)

PSI conference
9 June 2025, London

With thanks to Nicky Best (GSK), Olivier Collignon (GSK), Aaron Dane (Danestat), Nelson Kinnersley (Octa consulting), Satrajit Roychoudhury (Pfizer), Oliver Sailer (Boehringer-Ingelheim), Simon Wandel (Novartis), Maxine Ajimi (AstraZeneca), Beat Neuenschwander (Novartis), Olta Kodra (GSK) and the eSIG Historical Data

SARYGA
STATISTICS & METHODOLOGY

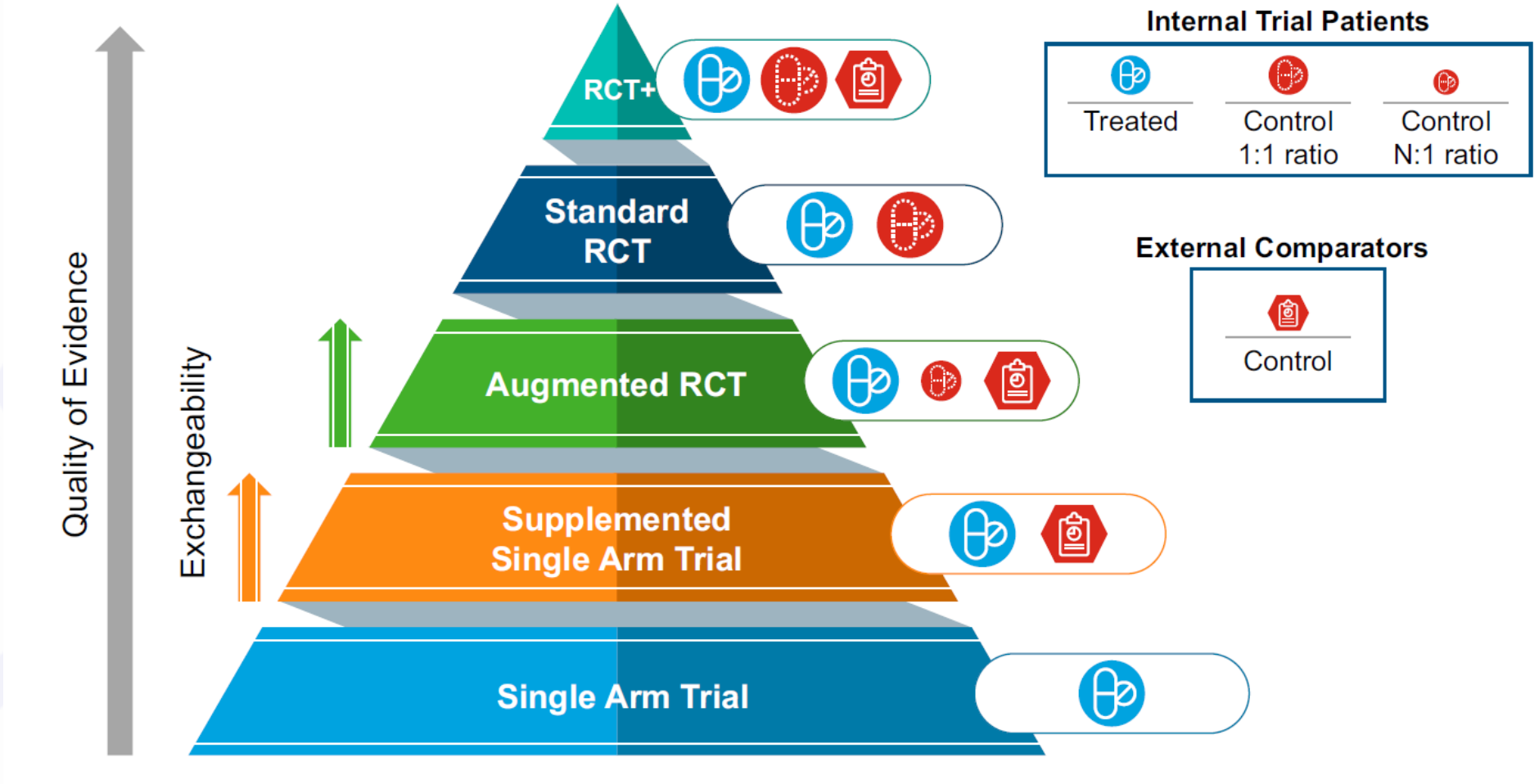
Acknowledgment and Disclaimer

- This presentation includes contributions from collaborative work with the European Special Interest Group (eSIG) on Historical Data
- In particular, some sections reflect proposals developed by the group as part of a Request for CHMP Qualification for Bayesian dynamic borrowing (on-going)

- Other parts of the presentation express my personal views and interpretations, and should not be understood as representing the position of the eSIG or its members

SARYGA
STATISTICS & METHODOLOGY

Quality of evidence



Gray et al (2020) - A Framework for Methodological Choice and Evidence Assessment for Studies Using External Comparators from Real-World Data

SARYGA
STATISTICS & METHODOLOGY

Framework for evaluation of a Bayesian Dynamic Borrowing (BDB) Design

Transparent, consistent approach

Part 1. Rationale for the chosen design and quality of external data*

Part 2. Comprehensive assessment of operating characteristics to evaluate trade-off between potential bias versus efficiency gain of the proposed BDB design

- **A. Operating characteristics of the chosen design:** expanded set of metrics including average type I error and pre-posterior probability of false positive result
- **B. Additional operating characteristics to support selection of design parameters:** summarise how operating characteristics vary with respect to key design parameters
- **C. Illustrative results of the chosen design under selected data scenarios:** understand how BDB design affects results for a range of potential observed study outcomes

Part 3. Reporting results and sensitivity analyses: including summaries of prior, data and posterior; assessment of concordance between external and current data and amount of information borrowed; robustness to key prior assumptions and impact of potential sources of bias

* Detailed discussion of the rationale for proposing a BDB design and the selection and quality if external data is highly specific to the disease and drug development context under consideration, and is beyond the scope of the request for qualification opinion

SARYGA
STATISTICS & METHODOLOGY

Framework for evaluation of a Bayesian Dynamic Borrowing (BDB) Design

Transparent, consistent approach

Part 1. Rationale for the chosen design and quality of external data*

- Context of use
- Motivation for borrowing external information
- Selection of historical data and their relevance to the target study population

Part 2. Cor
bias versu

A. Opera
and pre-p

B. Additi
operating

C. Illustrative results of the chosen design under selected data scenarios: understand how BDB design affects results for a range of potential observed study outcomes

Part 3. Reporting results and sensitivity analyses: including summaries of prior, data and posterior; assessment of concordance between external and current data and amount of information borrowed; robustness to key prior assumptions and impact of potential sources of bias

* Detailed discussion of the rationale for proposing a BDB design and the selection and quality if external data is highly specific to the disease and drug development context under consideration, and is beyond the scope of the request for qualification opinion

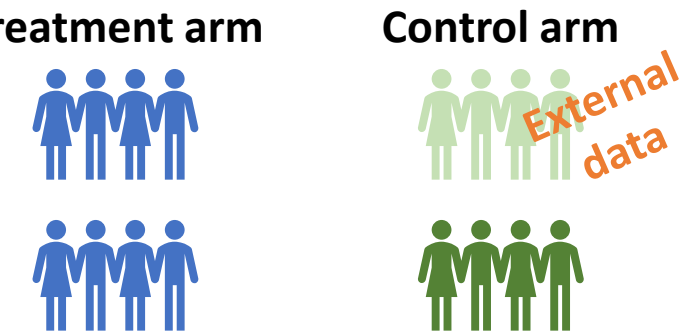
SARYGA
STATISTICS & METHODOLOGY

Context of Use: Favorable Clinical vs Statistical Contexts

Case 1: rare disease, unmet medical need, serious condition

Context

- **Rare disease**, serious condition
- Planned study design: **single arm** or with a **(small) control arm**
- External data available : **control arm data from 1 registry**
- Motivation for using external data: to “replace” study patients on the control arm by external data



Attractive clinical context

- **Very low feasibility** of recruiting enough patients
- **Ethical concerns and practical difficulties** in enrolling patients into a control arm
- High unmet medical need

Unfavorable statistical context

- **Limited and sparse external data**, not meeting clinical trial standards for data quality
- **Patient heterogeneity** is hard to quantify
- Only 1 external data source: no possibility to estimate **between-trial heterogeneity**

Clinical context	Medical need	High
	Ethical justification	High
	Likelihood of regulatory acceptance	High
Statistical context	Subjectivity in parameter definition	High
	Expected gains (precision, power)	Low
	Risk of bias in estimates	High

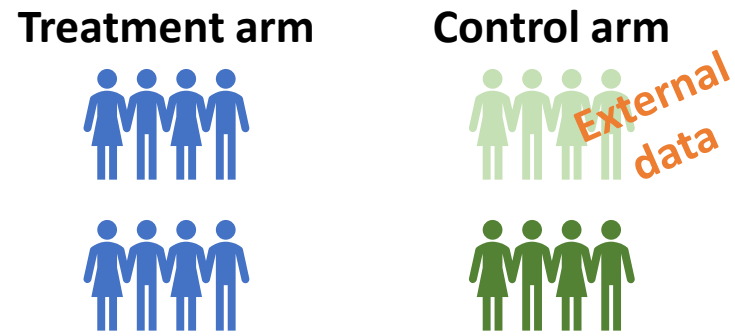
SARYGA
STATISTICS & METHODOLOGY

Context of Use: Favorable Clinical vs Statistical Contexts

Case 2: common disease, large population

Context

- Common, well-known disease with a **large patient population**
- Planned **regulatory** study: randomized, **placebo-controlled** clinical trial
- External data available: placebo arm data from **10 historical studies**
- Motivation for using external data: to “replace” study patients on the control arm by external data



Challenging clinical / regulatory context

- **High feasibility** of recruiting patients
- **Low likelihood of regulatory acceptance** for the use historical data in this setting
- However, potential cost reduction by using historical data and ethical advantage of reducing placebo exposure

Highly favorable statistical context

- **Large amount of high-quality** historical data
- **Patient heterogeneity** is well characterized and can be accounted for
- **Between-trial heterogeneity** can be estimated

Clinical context	Medical need	Low
	Ethical justification	High
	Likelihood of regulatory acceptance	Low
Statistical context	Subjectivity in parameter definition	Low
	Expected gains (precision, power)	High
	Risk of bias in estimates	Low

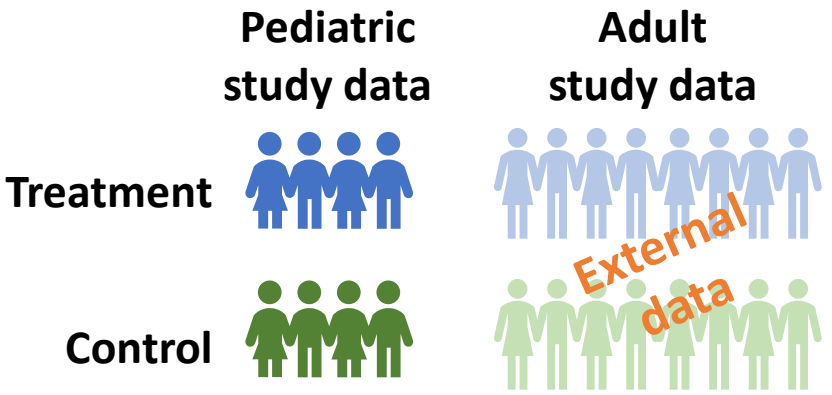
SARYGA
STATISTICS & METHODOLOGY

Context of Use: Favorable Clinical vs Statistical Contexts

Case 3: Pediatric extrapolation

Context

- Rare disease with a **drug already approved in adults**
- Planned **pediatric** study: randomized, **controlled** clinical trial
- External data available: one **Phase 3 study in adults**
- Motivation for using external data: supplement the planned pediatric sample size with adult data to improve efficiency



Attractive clinical context

- **Very low feasibility** of recruiting pediatric patients
- High unmet medical need in the pediatric population
- Good biological and clinical rationale for transposability from adults to children

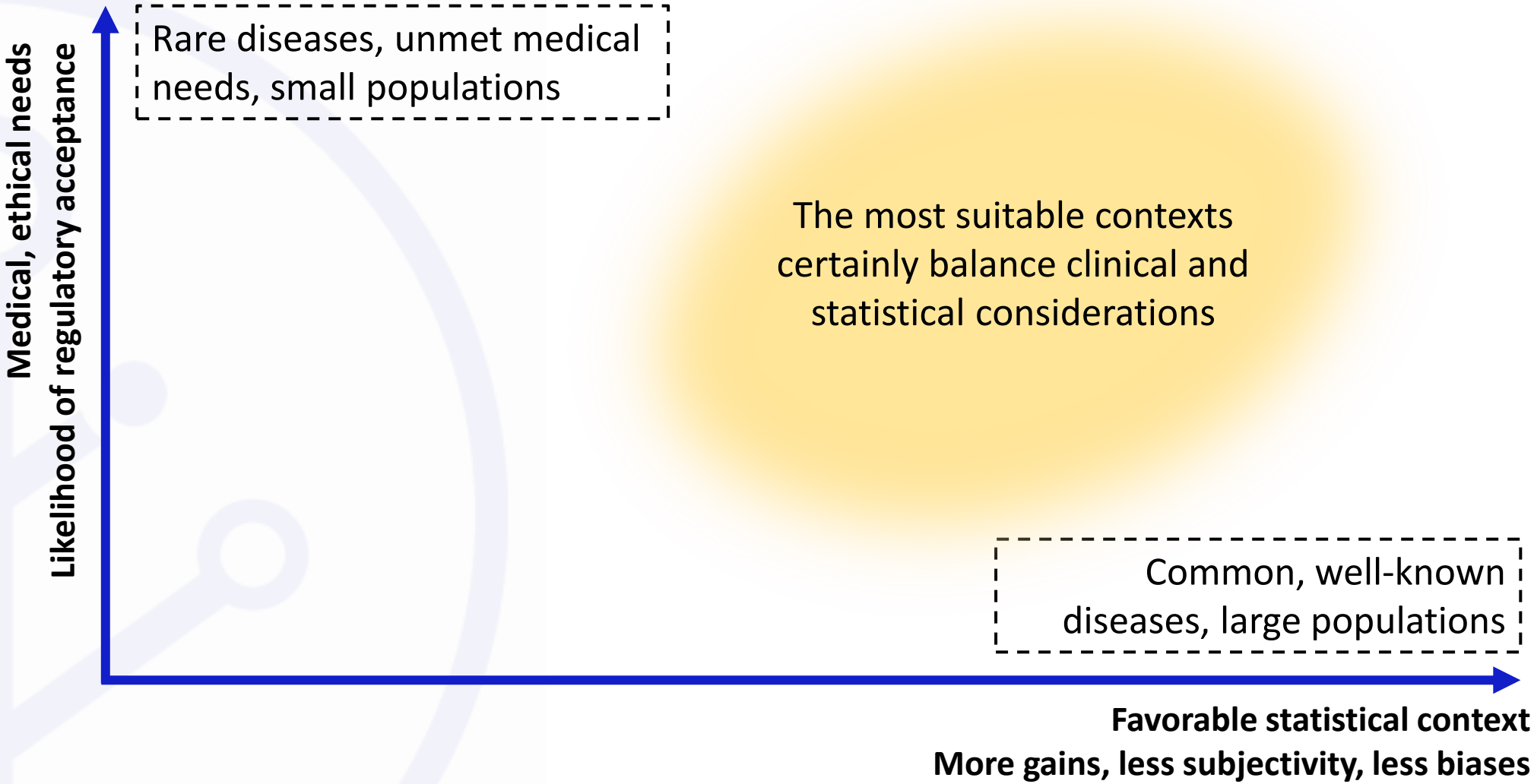
Relatively favorable statistical context

- **High-quality data from adults**, and individual patient data (IPD) available for covariate adjustment
- Only 1 historical data source: unable to estimate **between-trial heterogeneity**
- **Subjective assumption of transposability** from adults to children

Clinical context	Medical need	High
	Ethical justification	High
	Likelihood of regulatory acceptance	High
Statistical context	Subjectivity in parameter definition	High
	Expected gains (precision, power)	High
	Risk of bias in estimates	Mod.

SARYGA
STATISTICS & METHODOLOGY

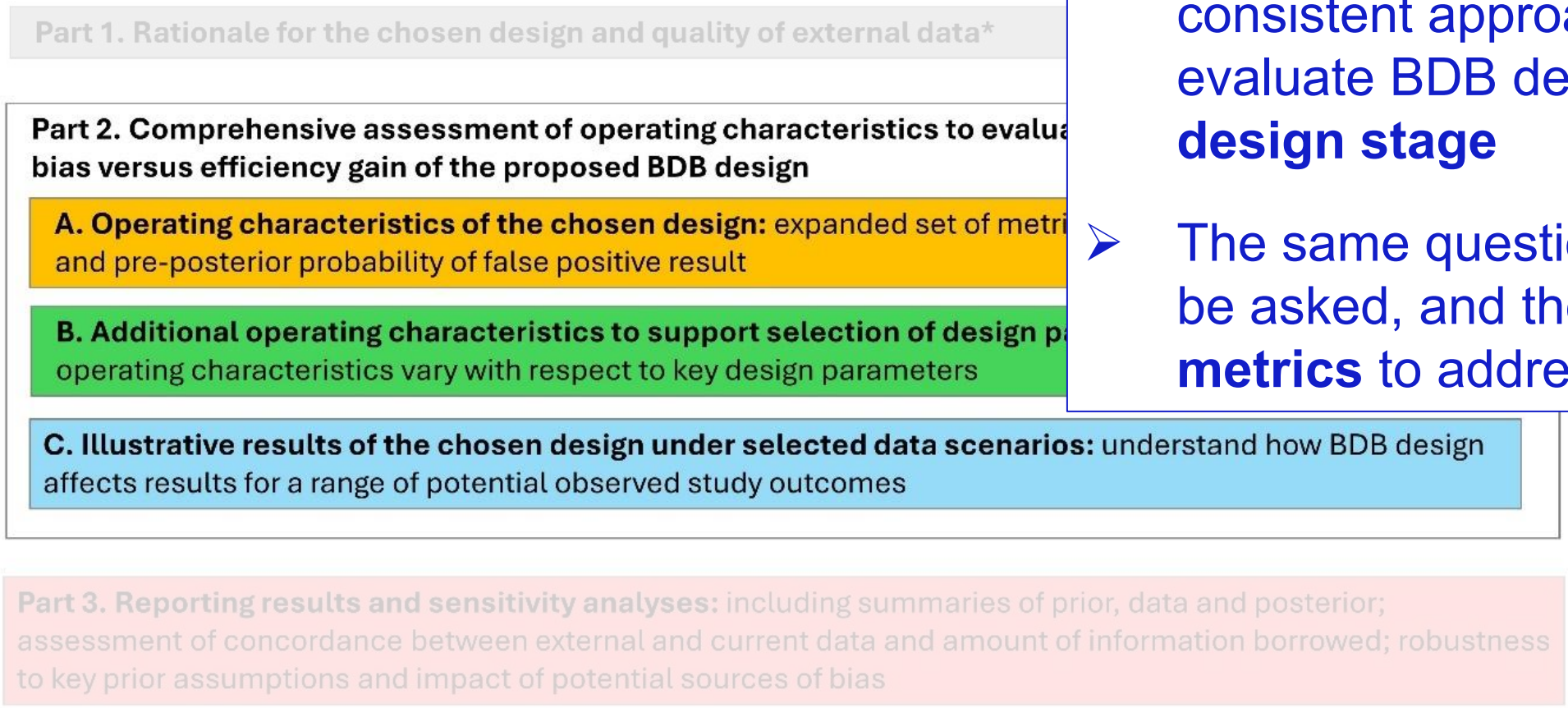
Context of Use: Favorable Clinical vs Statistical Contexts



SARYGA
STATISTICS & METHODOLOGY

Framework for evaluation of a Bayesian Dynamic Borrowing (BDB) Design

Transparent, consistent approach



- Whatever the context, a consistent approach exists to evaluate BDB designs **at the design stage**
- The same questions should be asked, and there are **metrics** to address them

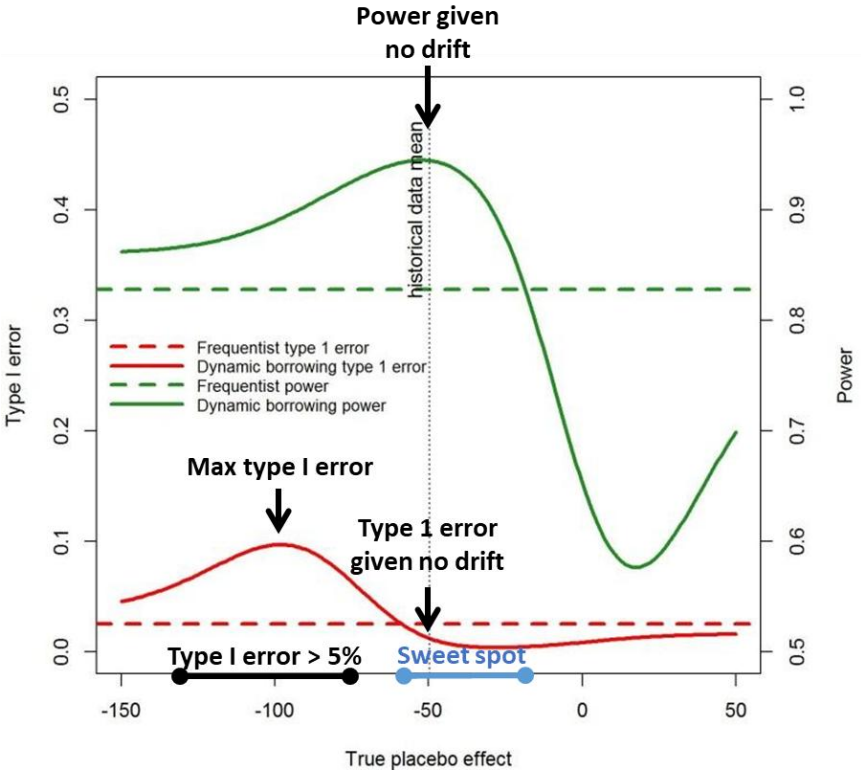
* Detailed discussion of the rationale for proposing a BDB design and the selection and quality if external data is highly specific to the disease and drug development context under consideration, and is beyond the scope of the request for qualification opinion

SARYGA
STATISTICS & METHODOLOGY

Key questions to evaluate the risk of false positive conclusion (type 1 error) in BDB designs

Borrowing on the control arm

- What is the type 1 error, if study and historical data are **perfectly aligned (no drift)**?
  → **Is the gain substantial?**
- What is the **maximum type 1 error** across all possible discrepencies between study and historical data?
  → **Is the increase acceptable?**
- What is the **average type 1 error** across plausible control effect values?
  - Given current knowledge of the control effect **(controlled*)**
  - Under skeptical assumptions
- Given current knowledge of the control effect, what is the **probability of a drift** between study and historical data?



* Best N, Ajimi M, Neuenschwander B, Saint-Hilary G, Wandel S (2024). Beyond the classical type I error: Bayesian metrics for Bayesian designs using informative priors. Statistics in Biopharmaceutical Research.

SARYGA
STATISTICS & METHODOLOGY

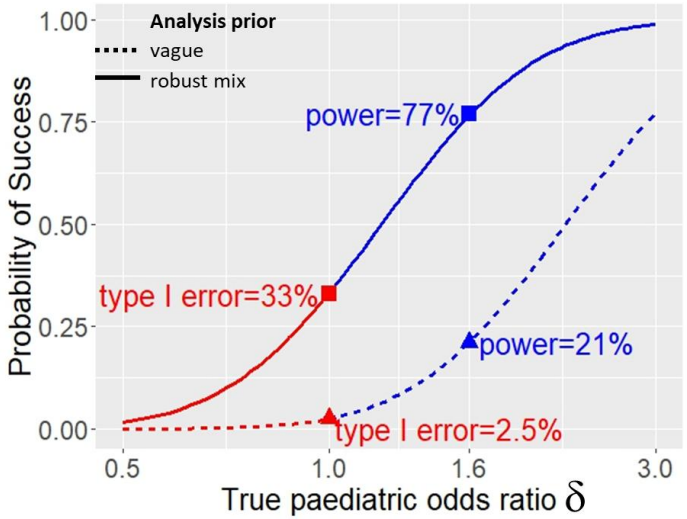
Key questions to evaluate the risk of false positive conclusion (type 1 error) in BDB designs

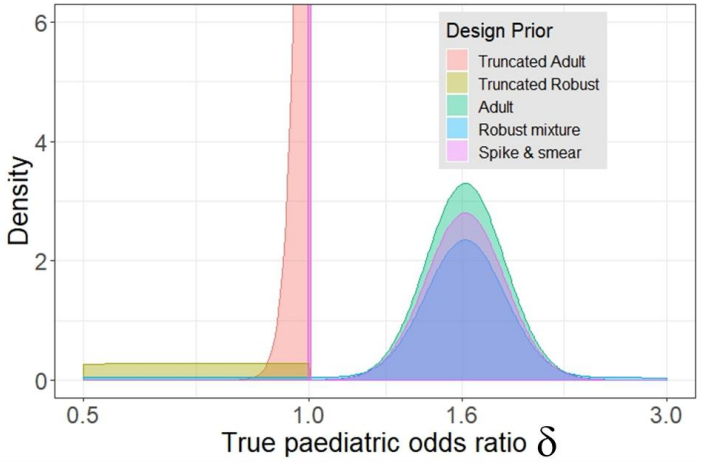
Borrowing on the treatment difference

Note: When borrowing on the treatment difference, the null hypothesis (no effect) inherently conflicts with historical data showing an effect, leading to Type I error inflation

- What is the type 1 error assuming there is no treatment effect ?
- What is the **probability of no treatment benefit** ?
  - Given current knowledge of the treatment effect
  - Under skeptical assumptions
- What is the **predictive probability of a false positive result** ? (Probability of no treatment benefit AND obtaining a false positive result)
  - Given current knowledge of the treatment effect
  - Under skeptical assumptions

→ **Is the risk of false positive result acceptable?**



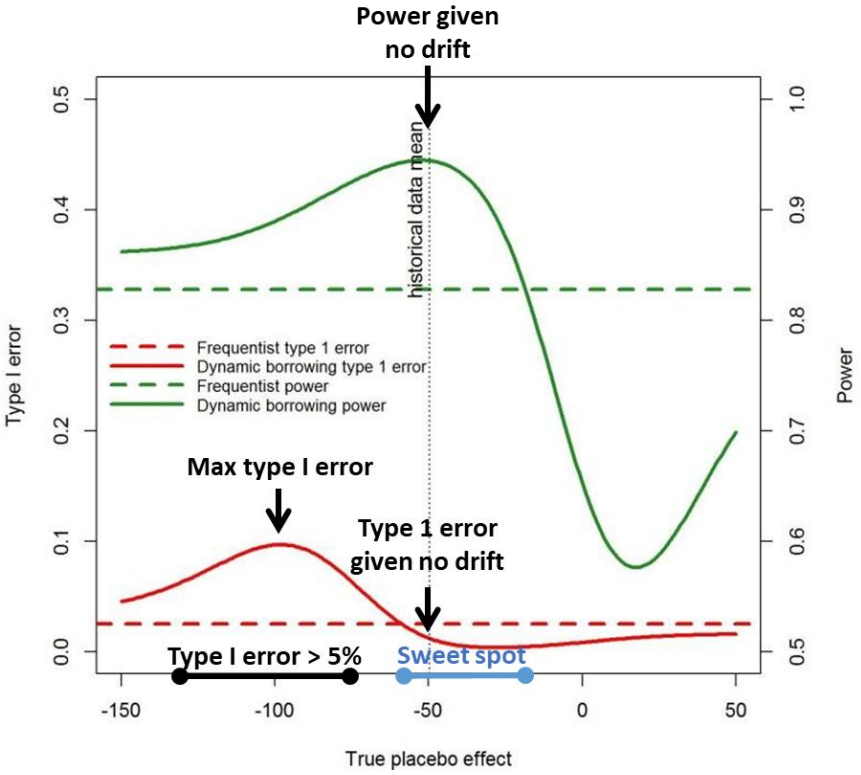


SARYGA
STATISTICS & METHODOLOGY

Key questions to evaluate the gains (power, MSE) with BDB designs compared to a classical frequentist design

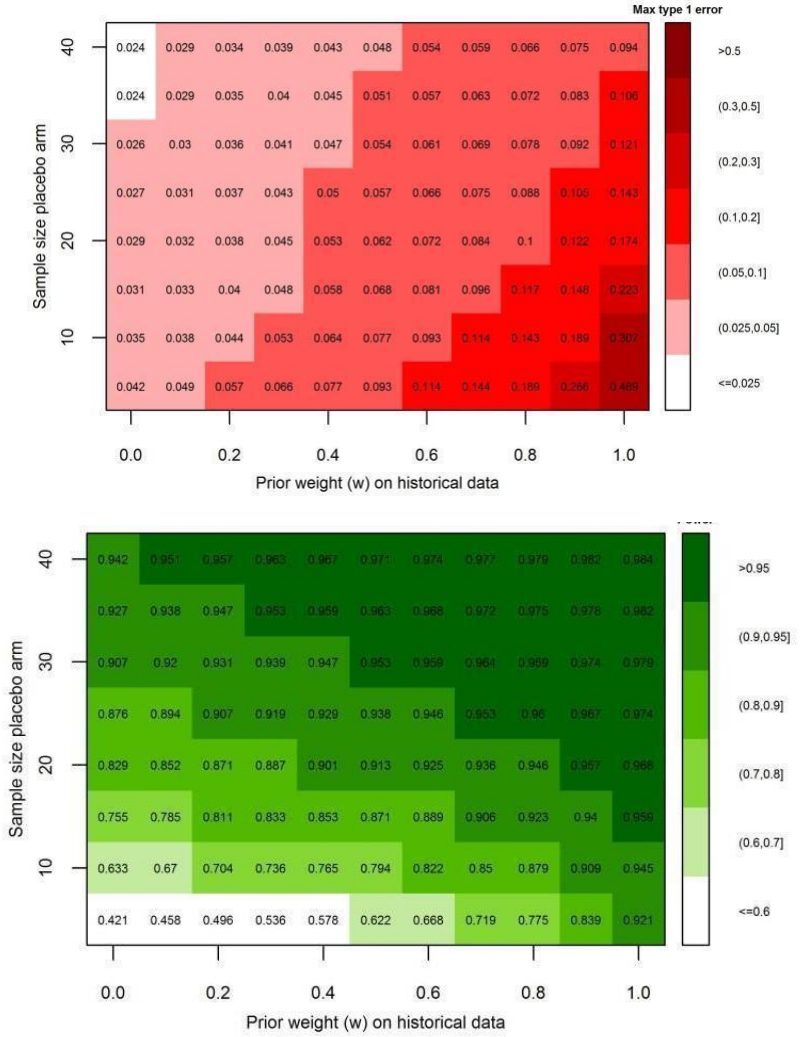
- What is the **power** of the BDB design?
  - Under realistic assumptions for the treatment effect
  - Averaged over plausible treatment effects based on current knowledge (i.e., **assurance**)
- What is the **probability of a "sweet spot"**?
  *(i.e., where the BDB design achieves both **higher power** and **lower Type I error** compared to the classical design)*
- What is the average **gain in precision**?
- What is the **bias in the point estimate**?
- What is the **mean squared error (MSE)**?

→ **Is there a reduction in MSE? Is it substantial?**



SARYGA
STATISTICS & METHODOLOGY

How to pre-specify the parameters?



Parameters to be specified at the design stage

- Variance of the vague components of the robust mixture prior
- Initial **weight** on the historical data (informative prior component)
- **Sample size** of the study

➢ What is the **optimal combination of weight and sample size** to provide good operating characteristics?

➢ What is the **prior effective sample size (ESS) of the informative prior**, i.e. the amount of information introduced by the historical data, expressed in terms of number of patients?

SARYGA
STATISTICS & METHODOLOGY

“What if”: Illustrative results under selected data scenarios

- What will the **study outcome be under different observed effects**?
  - If the observed effect is consistent with historical data
  - If the observed effect is more extreme than historical data (e.g., twice as large)
  - Etc.

Assumed observed placebo mean response value in new study	Proposed BDB design (w =80%)		Frequentist design
	Posterior weight ($\widetilde{w}$) on historically-informed evidence	Point estimate (posterior mean) of the placebo response [95% CrI]	Point estimate of the placebo response [95% CI]
-20	0.88	-36.6 [-56.1; -8.9]	-20 [-58.6; 18.6]
-50	**0.93**	-48.9 [-69.7;-30.4]	-50 [-88.6; -11.4]
-80	0.87	-63.5 [-93.2;-41.2]	-80 [-118.6; -41.4]
-110	0.56	-93.8 [-132.0;-57.3]	-110 [-148.6; -71.4]
-140	**0.19**	-132.6 [-165.9;-96.2]	-140 [-178.6;-101.4]

Consistent with historical data →

Prior-data conflict →

Observed placebo mean:

Worse *than in historical data*

Better *than in historical data*

SARYGA
STATISTICS & METHODOLOGY

Framework for evaluation of a Bayesian Dynamic Borrowing (BDB) Design

Transparent, consistent approach

Part 1. Rationale for the chosen design and quality of external data*

Part 2. Comprehensive assessment of operating characteristics to evaluate trade-off between potential bias versus efficiency gain of the proposed BDB design

A. **Operating characteristics of the chosen design:** expanded set of metrics including average type I error and pre-posterior probability of false positive result

B. **Additional operating characteristics to support sele**
operating characteristics vary with respect to key design p

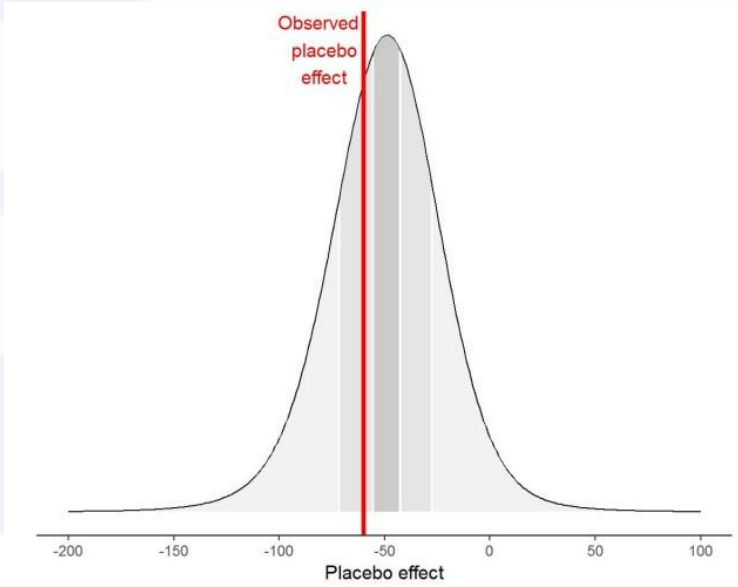
C. **Illustrative results of the chosen design under select**
affects results for a range of potential observed study outc

- At the analysis stage, upon obtaining the study results, **metrics** are available to evaluate the robustness of the conclusions

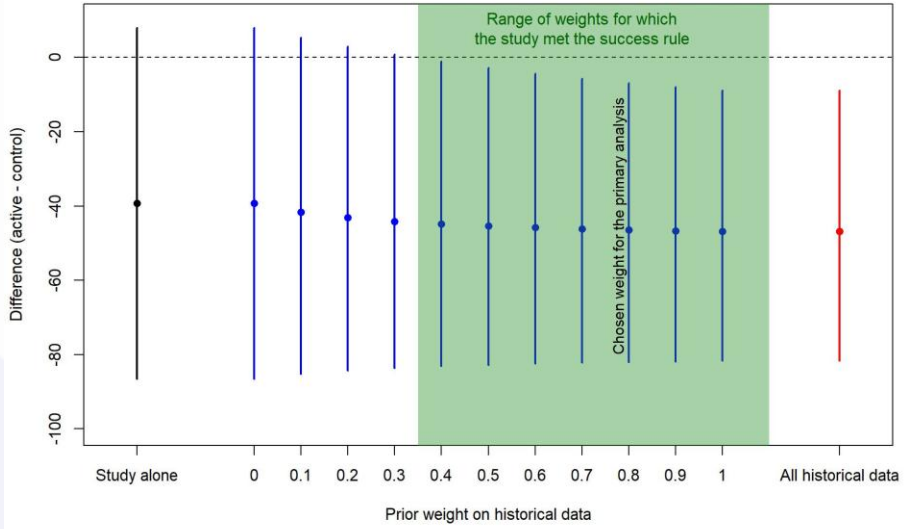
Part 3. Reporting results and sensitivity analyses: including summaries of prior, data and posterior; assessment of concordance between external and current data and amount of information borrowed; robustness to key prior assumptions and impact of potential sources of bias

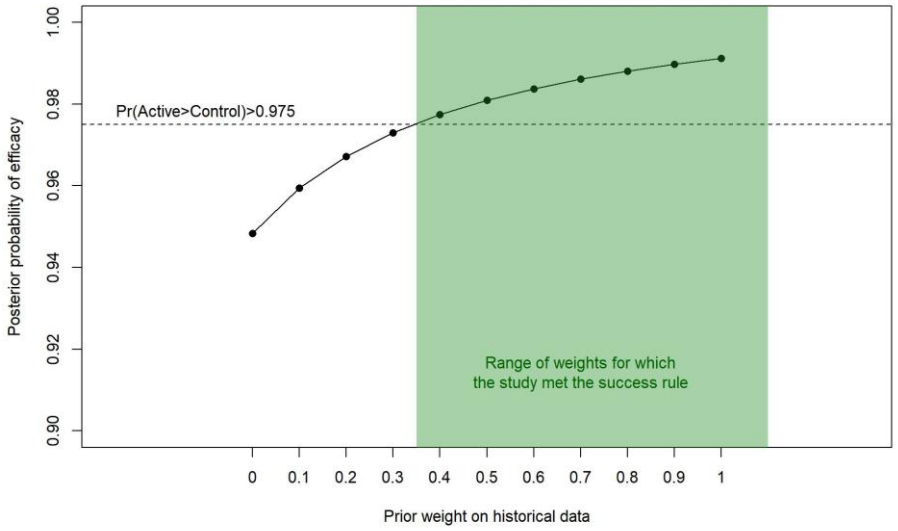
SARYGA
STATISTICS & METHODOLOGY

Upon obtaining the study results…



- What would the results have been **without borrowing from historical data**?
- Is the effect observed in the study **consistent with what was expected**, given current knowledge?
- Is the outcome of the study **sensitive to the prior weight** that was given to the historical data?





SARYGA
STATISTICS & METHODOLOGY

Discussion

- The **context of use** for BDB designs involves clinical, regulatory, and statistical considerations – which **may sometimes conflict**
- Once the context is justified and historical data are clearly identified, a structured framework is available to **fairly evaluate BDB designs**, both at the **design stage** and at the **analysis stage**
- Statistically, we are well equipped: multiple **metrics** now exist to assess performance, robustness, and risk
- The framework supports full **pre-specification** of design parameters, with scientific and technical rationales, and the ability to **quantify the impact of assumptions and subjectivity**

SARYGA
STATISTICS & METHODOLOGY