Adaptive Platform Trials
Complex and Innovative – but how useful?

Tobias Mielke, PhD
Senior Scientific Director, Quantitative Sciences Consulting
Janssen-Cilag GmbH – Statistics and Decision Sciences
April 19th 2023 – ADMTP Workshop Basel

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

Disclaimer

This presentation shares personal experiences and lessons learned on the design of platform trials. Thoughts expressed are my own and may not coincide with Janssen or IMI EU-PEARL position.

Experiences got a heavy push over the last 3 years – in particular also through discussions within EU-PEARL. Still, this is no EU-PEARL presentation (those come later).

Acknowledgements

Challenge, inspiration & motivation:

Mike Krams, Vlad Dragalin, Daniel Millar, Tony Vangeneugden

Introduction to adaptive designs:

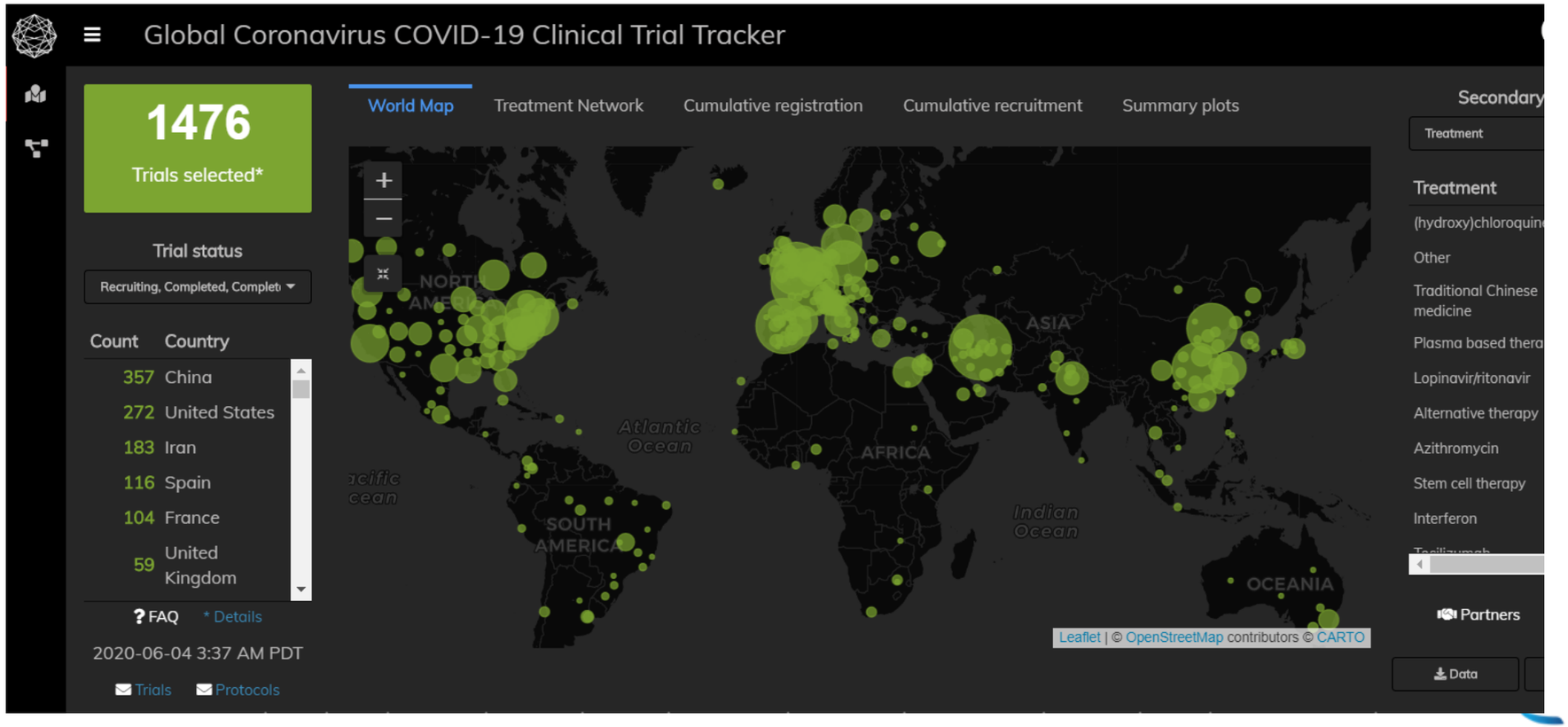
Silke Jörgens, Gernot Wassmer, Andy Grieve and Reinhard Eisebitt

Many platform discussions within EU-PEARL:

Tom Parke, Ursula Garczarek, Kyle Wathen & the team at Medical University Vienna

Motivation – 1: Covid 19

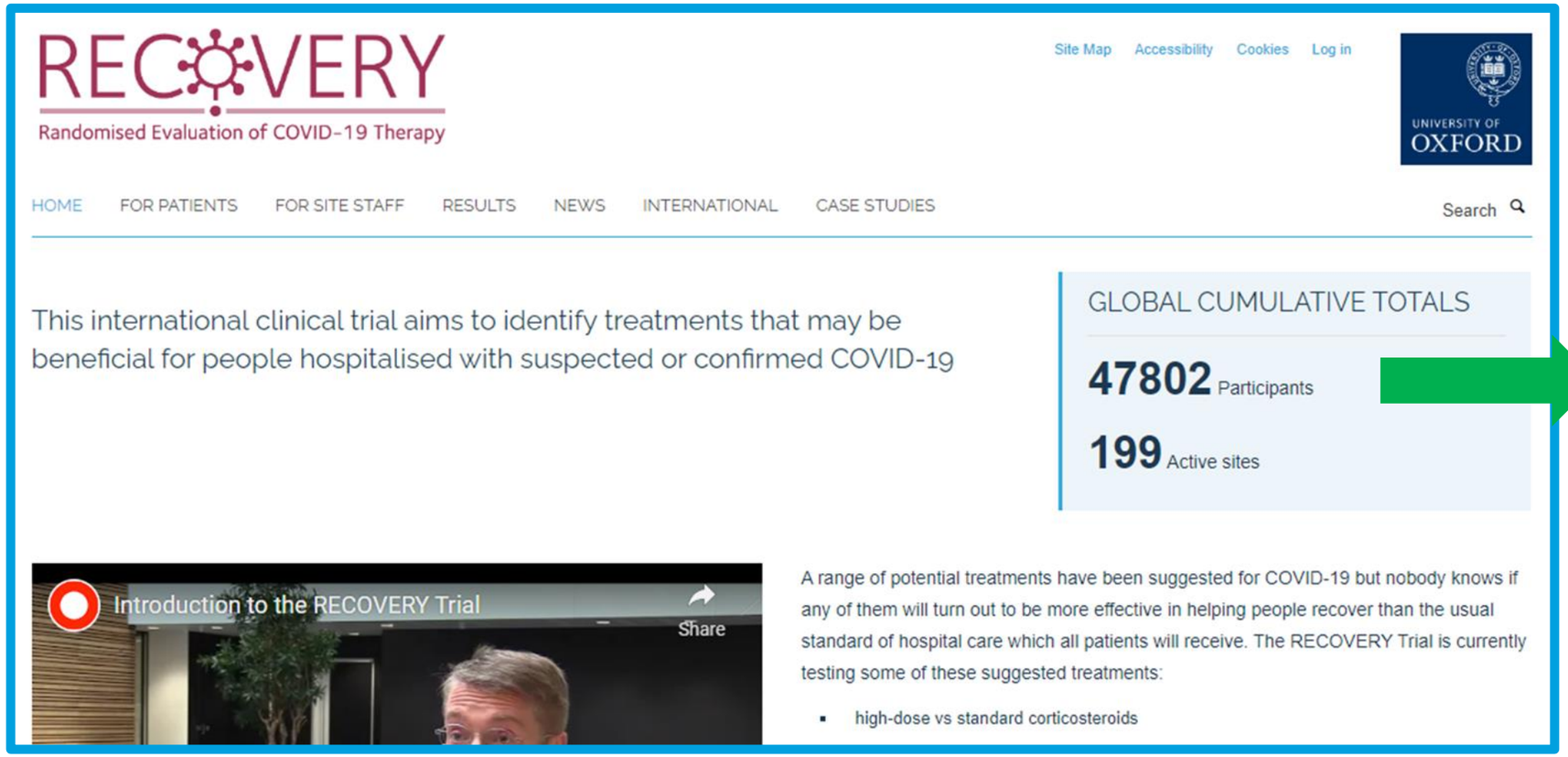
- Very many trials (~200 with HCQ) – many too small for definite answers
- Randomization? Common protocols? Endpoint definitions?



Covid19-trials.com by Cytel – assessed in June 2020

Motivation – 1: Covid 19

- One common indication centric platform trial evaluating many treatment options
- A common "platform" for patients, investigators and possibly intervention owners



Many…
- arms tested
- decisions supported
- patients saved

Recoverytrial.net – assessed in July 2022

Motivation – 2: Rare diseases

The opposite of Covid-19, but similar problem…

Assume 3 candidate interventions and 100 patients / year

3 separate trials

N=100

Drug A ~16	Ctrl ~16
Drug B ~16	Ctrl ~16
Drug C ~16	Ctrl ~16

1 multi-armed trial

N=100

Drug A ~25	
Drug B ~25	Ctrl ~25
Drug C ~25	

More data & Higher chance to get novel drug & Larger awareness

Motivation – 3: Oncology – I-SPY 2

*Platforms can be **quite complex***

Implementing Bayesian methods:

1. Randomization based on tumor subtype
2. Tumor assessment
3. Statistical modelling to update "response rate" (here predictive probability that drug will succeed in Ph3)
4. Stop/Graduate/Continue
5. Update randomization

No multiplicity/adaptive design control

Flexible size [20,120]

Details in appendix to Park et al. (2016)

https://www.ispytrials.org/i-spy-platform/i-spy2

https://www.nejm.org/doi/suppl/10.1056/NEJMoa1513750/suppl_file/nejmoa1513750_protocol.pdf

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Motivation – 3: Oncology – STAMPEDE

*Platforms could also be **relatively** simple*

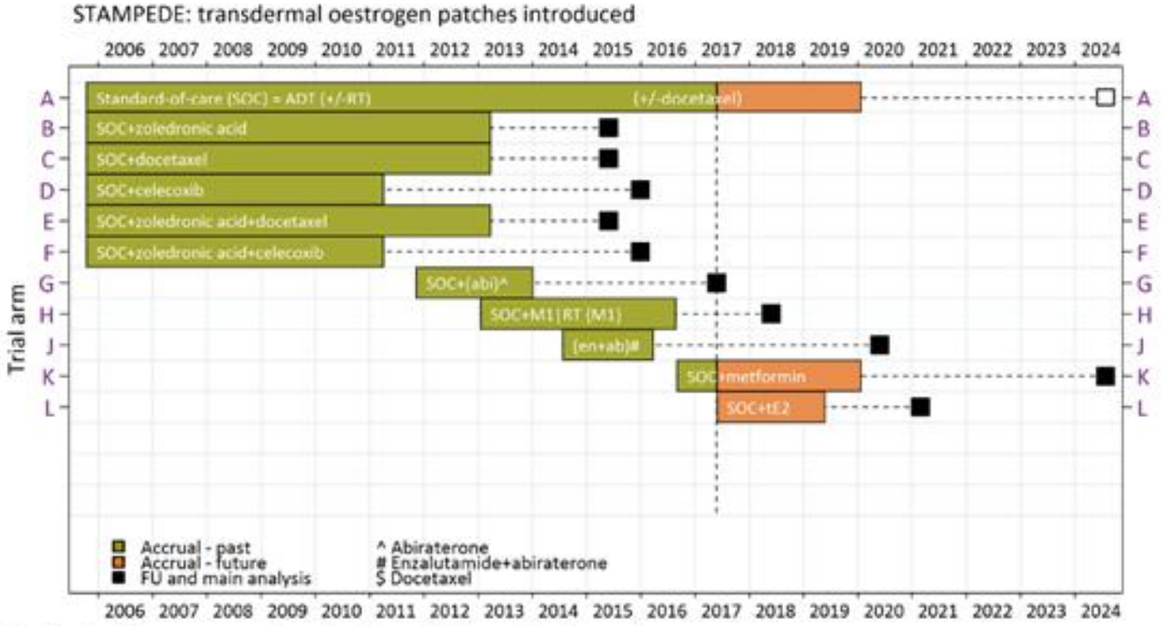
Objective: each arm vs. ctrl

Stages:

1. Pilot/feasibility/safety phase
2. Activity stage (FFS, futility)
3. Efficacy stage (OS)

Monitoring:

- Events on (contemporaneous) control



Note: dotted line represents activation of this protocol version

Notes:

- Started as MAMS in 2005, changed to “Platform” in 2011
- Very valuable open publications on operations and statistics

http://www.stampedetrial.org/centres/information-on-stampede/

Platform trials are always complex*…

Study populations
Biomarker driven allocation?

Statistical modelling
Frequentist vs. Bayesian?

Allocation rules
Fixed, RAR, dynamic

Data sharing
All, concurrent, modelling?

Decision making
Size, arms, populations…

Sample size
Fixed, adaptive, perpetual

Endpoints
Primary, secondary, surrogates…

Multiplicity
No control, strict, dynamic?

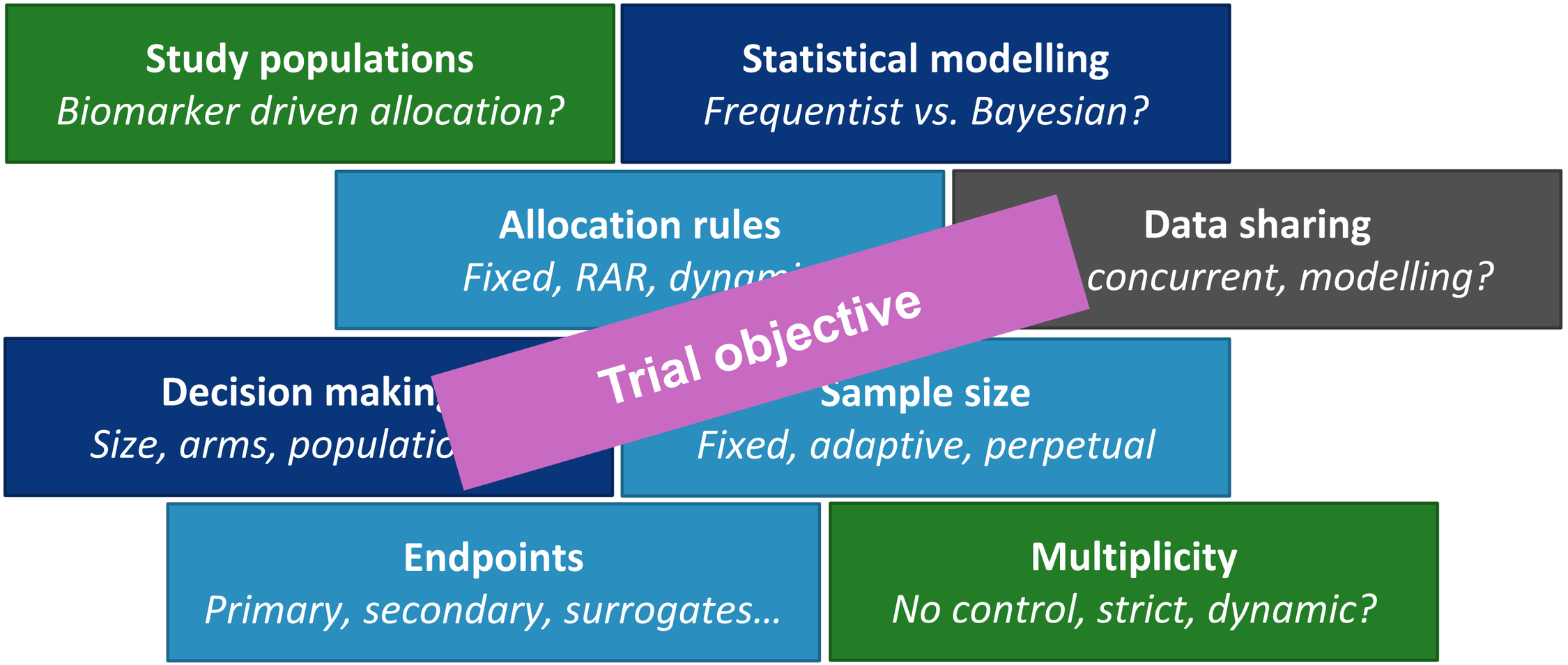
Today's presentation: Some experiences & things to consider

More details: Next session

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

* *at least to me*

Platform trials are always complex*…



Today's presentation: Some experiences & things to consider

More details: Next session

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

* *at least to me*

How to do research efficiently?

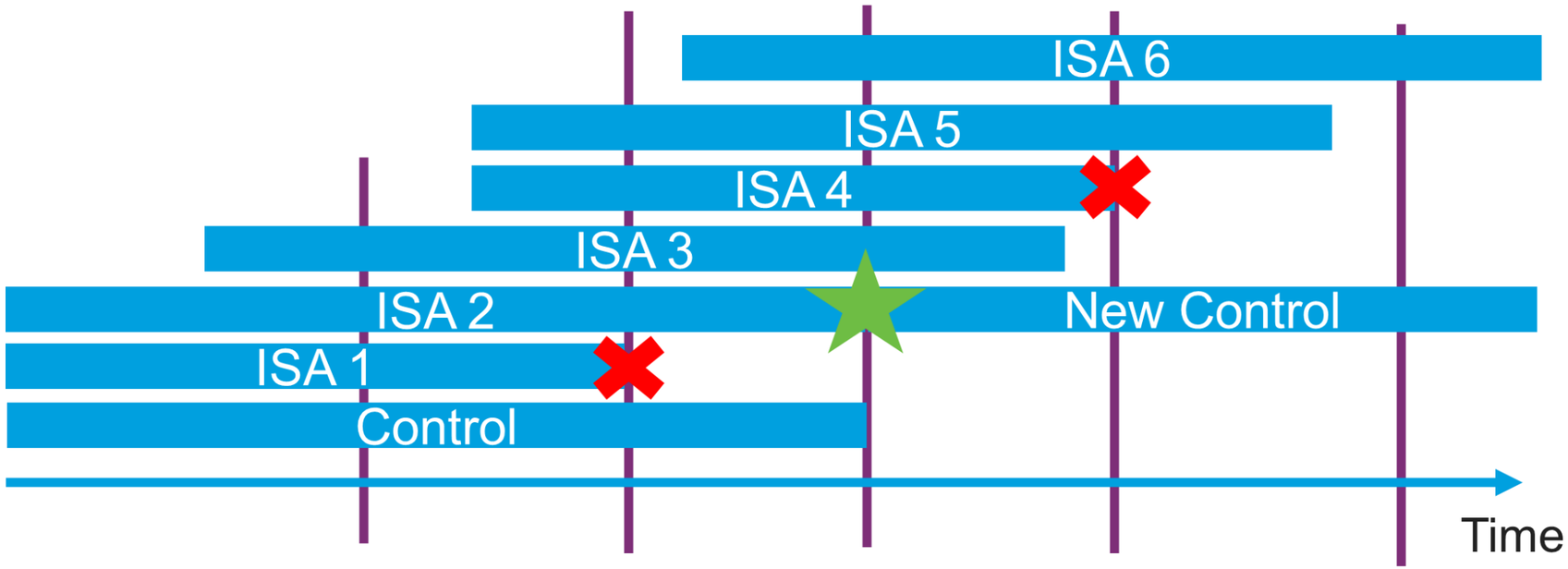
Design	Allocation	Total size	Per active arm	Control size (Total)	Power (individual)	FWER (not adj.)	Interventions tested
Standard	1:1	800	100	400	90.2%	9.6%	4
Shared control	1:4	800	100	400	98.4%	9.5%	4
Square root	1:2	600	100	200	96.4%	9.1%	4
SR + 90%	1:2	450	75	150	90.2%	9.1%	4
N=800 + 90% +SR	1:3	800	66.67	200	90.2%	19.2%	9

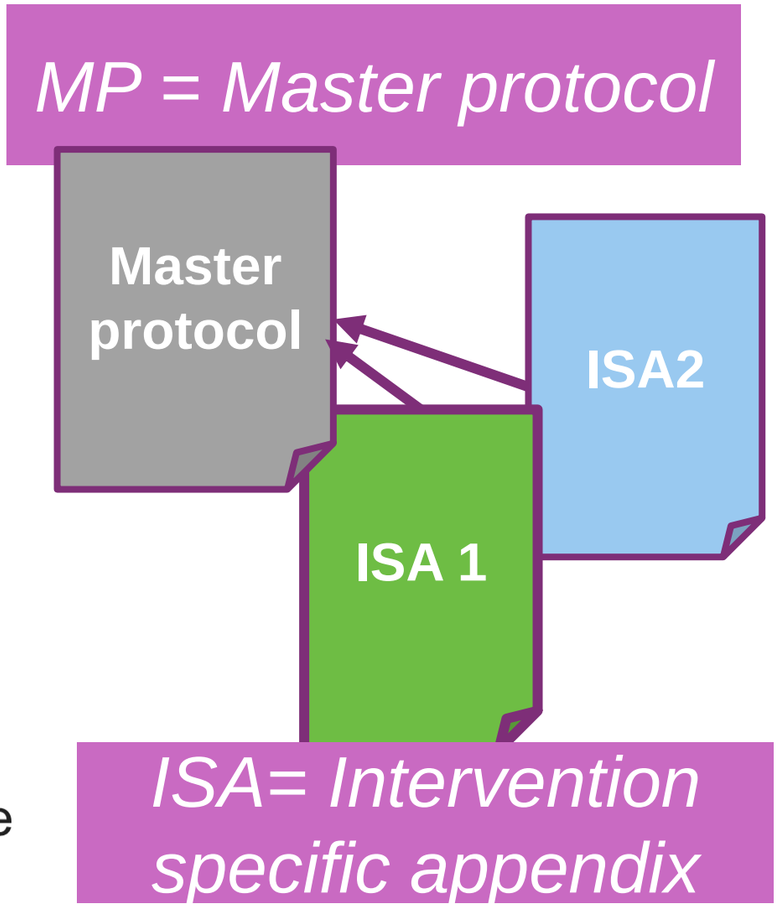
- FWER = 19.2% => **Huge problem, if we don't care with separate trials?**
- N=200 vs. N=450 => recruitment time might severely increase: **Awareness?**
- You rarely have 9 interventions to be tested simultaneously: **Allow for delayed start**
- Not all arms equally effective: **Drop arms early?**

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

The solution: Platform trials

Platform trials:





Solving some problems, while adding many more complexities

How innovative are platform trials?

When is something called innovative?

- Early Phase screening designs? → 90ies (e.g. Yao et al. (1996))*
- Multi-armed adaptive designs? → Early 2000 (e.g. Bretz et al. (2006))
- Response adaptive designs? → 90ies (e.g. Krams et al. (2003, ASTIN))
- STAMPEDE: Started in 2005
- Number of publications and implementations: Quite large.

… still: Many additional things require consideration

** Most likely there are some publications by Fisher advocating platform type designs*

How do I design a platform trial?

Too many design parameters…

Study populations
Biomarker driven allocation?

Statistical modelling
Frequentist vs. Bayesian?

Allocation rules
Fixed, RAR, dynamic

Data sharing
All, concurrent, modelling?

Decision making
Size, arms, populations…

Sample size
Fixed, adaptive, perpetual

Endpoints
Primary, secondary, surrogates…

Multiplicity
No control, strict, dynamic?

… depending on too many assumptions

Recruitment rate

Effect assumptions

Timing of entry of new arms

Decisions on other arms

Subgroup effects

Easy to get lost → start simple: **What is my problem now?**

- **My only interest: assess intervention X vs. Ctrl**
- Design your ISA for this. Then combine & adjust.

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

How to design a platform trial?

The EU-PEARL Process

Starting design

Too slow?

Sprinkle 1
Early stopping

Too low power?

Sprinkle 2
Sharing control data

Too many participants on ineffective arms?

Final design

Simulate, discuss, adjust

Sprinkle 3
Response adaptive randomization

Too late decisions

...

Sprinkle 4 Advanced modelling

Sprinkle 5,6,7...

How to design a platform trial?

Multiplicity may not be the deal breaker

The EU-PEARL Process

Starting design

Too slow?

Too low power?

Too many participants on ineffective arms?

Too late decisions

...

Simulate, discuss, adjust

Sprinkle 1 Early stopping

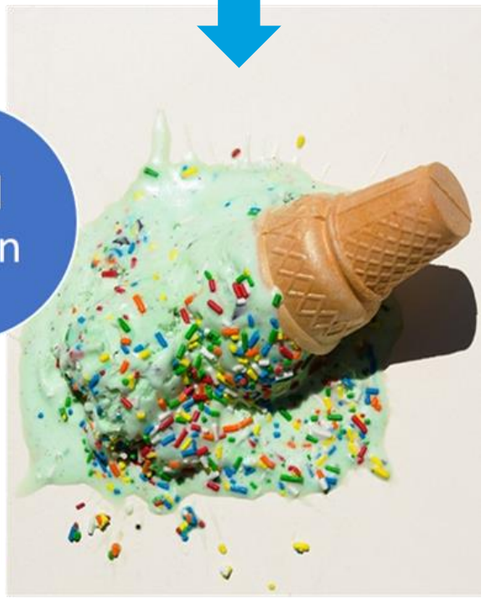
Sprinkle 2 Sharing control data

Sprinkle 3 Response adaptive randomization

Sprinkle 4 Advanced modelling

Sprinkle 5,6,7...

Final design



How to design a platform trial?

What you need are

Simulations

… to enable discussions.

Senior decision makers don’t care about the error rate, if

- trial takes too long
- is too expensive or
- does not allow for internal decision making

Problem: Also simulations for platform trials are complex*

* e.g. Meyer et al. 2021

Case 1: Internal Covid-19 Platform

"We're now asked to set up (urgently) a platform trial in moderate COVID-19 patients. Can you also help with this trial?

Major stats questions:

- *Relevant primary endpoint*
- *Sample size*
- *Control arm size for subsequent ISA"*

Highly relevant + Urgent + Adaptive/Platform

A most interesting challenge to join!

Case 1: Internal Covid-19 Platform

Rollercoaster of study design options

Type of design:

- No platform, but seamless Phase II/III
- No platform, but PoC only
- Platform, but Phase II only
- Platform, with ISA powered for EUA

Study population:

- Inpatient vs. outpatient
- High-risk vs. all-comers? Early-comers vs. all-comers?
- Endpoint?

Similar discussions elsewhere, e.g. PRINCIPLE Trial presentation by Berry Consultants

Case 1: Internal Covid-19 Platform

One week after initiation: Internal brainstorm session

Discussion points:

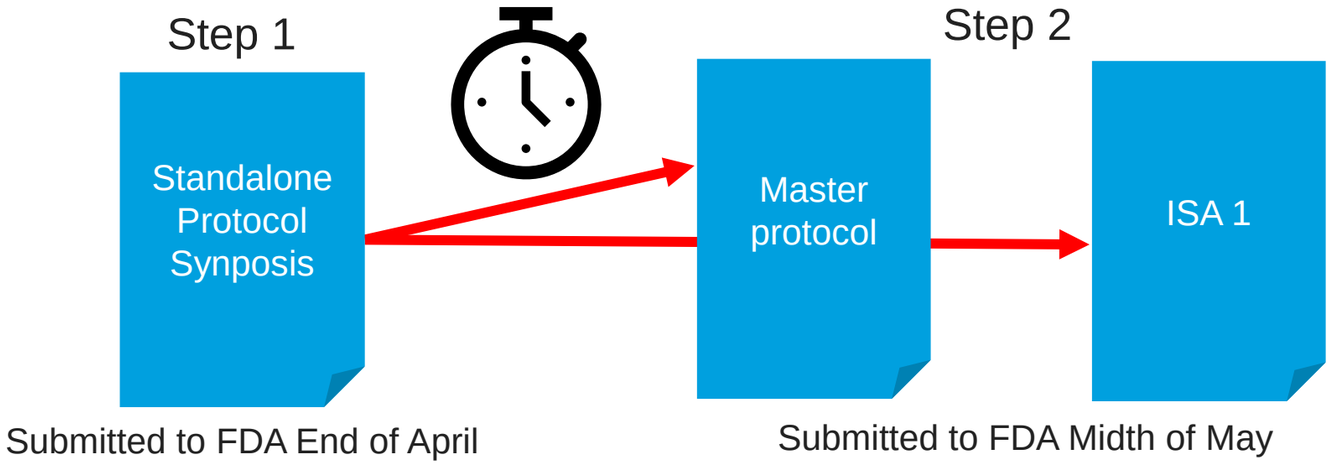
- What is an appropriate endpoint? Allow for adaptive change?
- What flexibility to allow?
- Multiplicity: Need to correct?
- How to randomize across ISAs?
  - Response adaptive?
  - 3 tiered approach – “Confirm” vs. “Screening” vs. “Pause”?

Can this be operationalized at all with fast recruitment?

Case 1: Internal Covid-19 Platform

Approach to the Platform Design Submission

Co-development of first ISA and platform:



- Agreement on ISA1 design (Top priority)
  - Study/ISA could stand on it‘s own *(Adaptive PE design fit for EUA)*
- Agreement on master protocol (Priority 2)
  - Everything set up if more compounds need investigation

Case 1: Internal Covid-19 Platform

Resulting Protocols

Lessons learned:

1. Master protocol development can go very fast – if required.
2. Uncertainties on some design elements may derail discussions
3. Maintain focus on the individual substudy objectives and needs
4. ... but enable flexibility for the future (not too rigid MP)

→ **This project:** Closed prior to FPI

Case 2: Internal Confirmatory Platform

Outcome from a Brainstorm session in different indication

"Given the number of potential combinations of intervention candidates of interest, a registrational platform trial should be considered."

Identified opportunity:

Sharing of control data → Decreased total sample size

Disclaimer: Information on the indication may not be provided.

Case 2: Internal Confirmatory Platform

Approach 1 – Fall 2020

Platform requires hand-shake between functions:

- Held separate high-level discussions with individual (non-stats) functions
- Aim: get individual alignment

Consensus meeting:

- Large concerns from clinical
- All functions followed clinical

… the ship left the harbor - no chance to adjust trajectory

Case 2: Internal Confirmatory Platform

Approach 2 – Spring-Summer 2021

Lesson learned: Increase preparedness

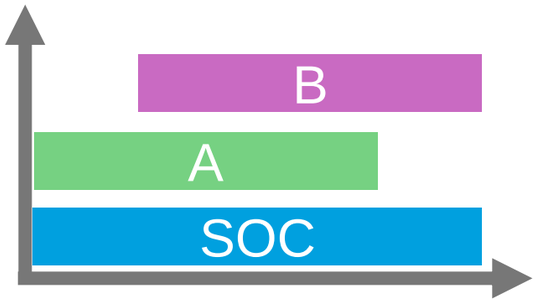
- Workshop series with project teams
- **Target outcome:** Co-development of a platform trial

*Discussion rule: Not "Yes/No", but **"How?"***

- Kickoff: Presentation of the potential opportunity
- Session 1: Past lessons learned: Stats, Operations & Regulatory
- Session 2: Address challenges: **Regulatory**, Operations, Organizational

Case 2: Internal Confirmatory Platform

Approach 2 – Spring-Summer 2021



Statistical outcomes presented:

- Decrease in total sample size
- Longer duration due to competitive enrollment (assuming 47/m)

Ctrl (mOS=12 months)*	A		B		Total
Assumptions	HR=75%, N=1200, 516 events		HR=75%, N=1200, 516 events		Sample size (per comp)
Design characteristic	Power	Duration	Power	Duration	
Separate studies	~90%	**23**	~90%	23	1071
Platform with 2 ISAs (same start)	~90%	29	~90%	29	716
Platform with 2 ISAs (+1 year)	~90%	24	~90%	**26**	927
Platform with 3 ISAs (yearly entry)	~90%	24	~90%	29, 27	849
Platform with 5 ISAs (yearly entry)	~90%	24	~90%	29, 30, 30, 28	807

*Numbers based on Mok et al. (2019), Keynote-042

How useful was this platform?

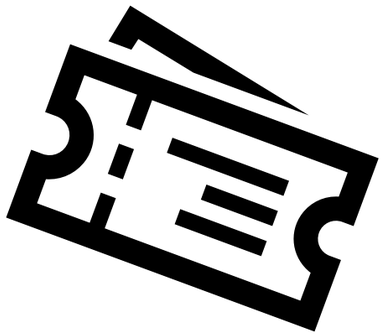
Concerns observed:

- Statistical efficiency? Dependent on overlap in recruitment
- Time savings? Depends on perspective…
- Multiplicity? Addressing for it → Lowers benefit.
- Regulatory and operational risk? **Increased**
- Time to study start? **Increased**

Case 2: Internal Confirmatory Platform

Approach 2 – Spring-Summer 2021

Session 2 outcome



"You need to buy a ticket to win the lottery"

Action: submit a question on the considered platform for scientific advice

Case 2: Internal Confirmatory Platform

Approach 2 – Spring-Summer 2021

Submitted question on very simplistic platform design:

- Time to event endpoint.
- Independent research hypotheses => No multiplicity adjustment
- Stable SoC arm and objective response assessment => Non-concurrent controls (unless strong difference observed)
- ISA designs with interim analyses for early success per ISA.
- Allocation ratio: to be defined by blinded study team.

Case 2: Internal Confirmatory Platform

Approach 2 – Mixed – but very valuable feedback

Particular comments:

- Control arm (SoC) and potential change of control
- Some openness to non-concurrent controls - need appropriate model
- Type-1 error control with respect to **multiple treatment arms** against shared control in **multiple efficacy analyses** and **change of allocation ratio required**.
- Adding new arms may require change of sample size of existing arms (e.g. # of events).
- Many concerns on study population

Complexities?

We tried to make the platform design very simple.

Still, from statistical perspective:

Platform trials are complex

Even if you want to make them non-complex.

Why?

"Independent decision making" may not hold, even if you plan for it

Complex? Allocation ratios

Why there is no “fixed allocation” in platform trials

	Period 1	Period 2	Period 3	Total
# of interventions	3	2	1	
Total size	200	100	50	350
Square root	36.6%	41.4%	50.0%	39.9%
	1:1.73	1:1.41	1:1.00	
Per arm (1:1)	25.0%	33.3%	50.0%	30.9%
	1:1	1:1	1:1	
Total fixed (2:1)	33.3%	33.3%	33.3%	33.3%
	1:1.5	1:1	1:0.5	

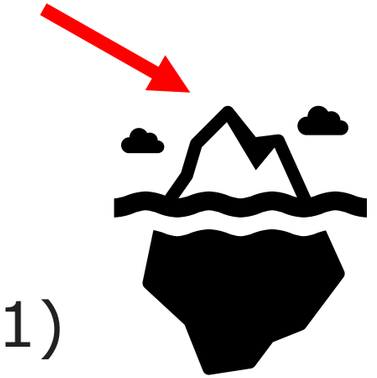
What is the implication on the power/required sample size?

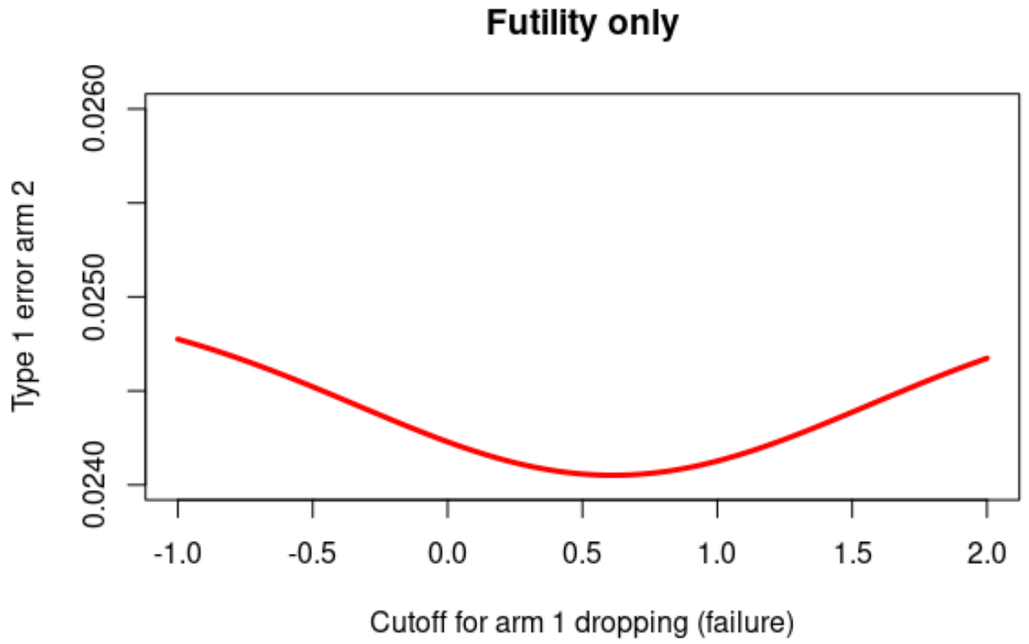
Always implement re-estimation? Is this blinded?

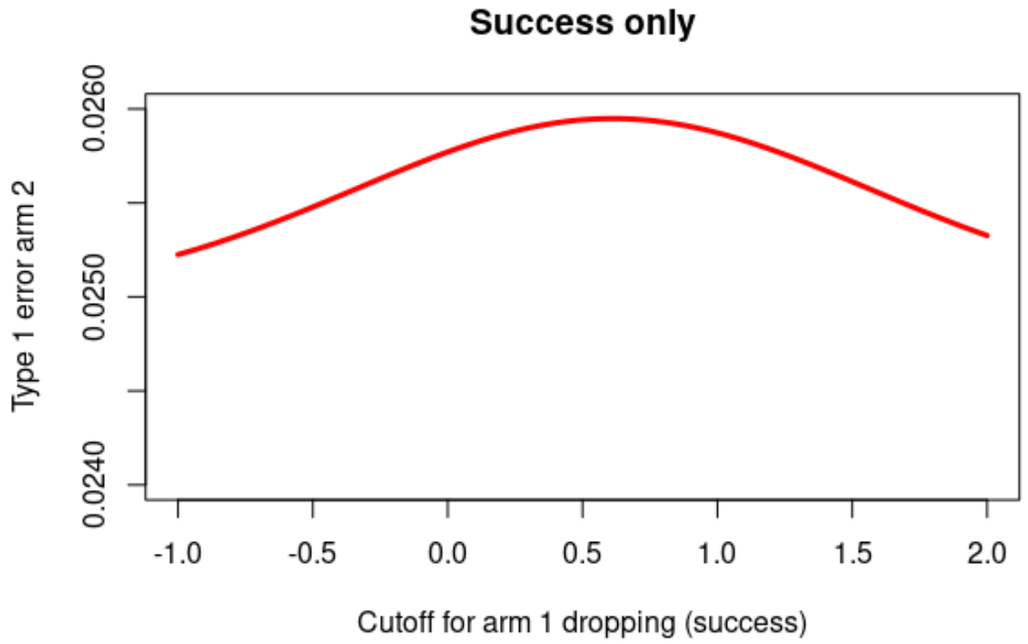
Complex? Interim Decisions

How interim decisions on one arm may impact the next arm…

- Assume square-root allocation rule and two arms
- At IA: decision on arm 1 vs. control for futility/success.
- Arm 2 to continue with updated allocation ratio (1:1.41 or 1:1)







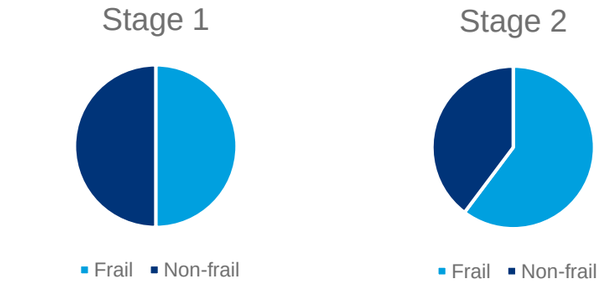
Complex? Populations (1)

How addition of an arm may bias the analysis…

- New intervention not allowed in “frail participants”
- … smaller potential advantage vs. SOC is “non-frail participants”

Simplistic design and analysis:

	Stage 1, N=200	Stage 2, N=200	Total
Frail – Control	50	50	100
Non-frail Control	50	33	83
Frail – TRT A	50	50	100
Non-frail TRT A	50	33	83
Frail – TRT B	-	0	0
Non-frail TRT B	-	33	33



Biased estimate of treatment effect, if not stratified

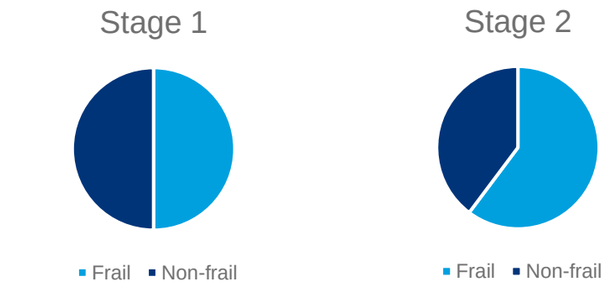
Complex? Populations (2)

How decision on one arm may impact second arm…

- Interim analysis for **B** suggests potential advantage vs. SOC is "non-frail participants"

Simplistic design and analysis:

	Stage 1, N=200	Stage 2, N=200	Total
Frail – Control	33	50	83
Non-frail Control	33	33	66
Frail – TRT A	33	50	83
Non-frail TRT A	33	33	66
Frail – TRT B	33	0	33
Non-frail TRT B	33	33	66



Enrichment in B results in de-richment in A
50% → 44%

Complex? Adding arms

Multiplicity strategies for “number of failed arms”

“Online error-rate control”*

- Reason for application: Potential inflation in FWER.
- **Simplistic:** Recycle alpha for successes, while losing alpha for non-successful treatments.

When entering the trial today, why should I pay penalty for an intervention failing 5 years ago?

… an incentive against platform trials

*E.g. Robertson et al. 2023

How complex are platform trials?

Very complex

- Implications of allocation rule on power and size?
- How to integrate non-concurrent control data?
- How may decisions on some arms impact other arms?
- What is the **optimal design** – addressing all uncertainties?
- … and how to execute those flawlessly?

How useful?

Statistical Criteria

Problem: “How to measure the utility of a platform trial?”

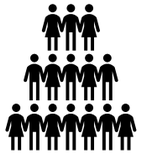
Number of treatment successes?

Time to first success

Average number of significant results?

What is most relevant?

Total sample size



Number of tested treatments?

Power for intervention X

Case 3: EU-PEARL Tuberculosis Platform

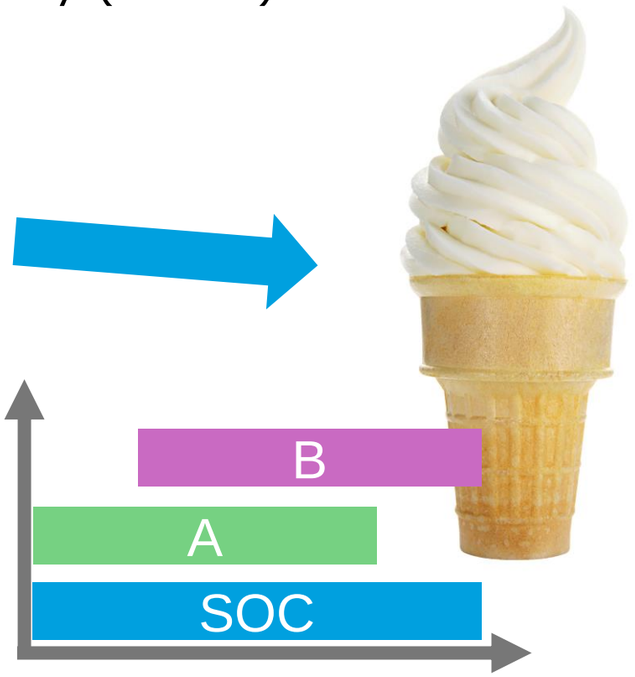
Research Objective: Simplify or shorten regimen without losing on efficacy (DS-TB)

Design motivated by:

- STEP design *(Phillips, P.P.J. et al.,* BMC Med, 2016;14(51)*)*
- PanACEA MAMS trials *(Phillips, P.P.J. et al.,* J Infect Dis., 2012;205*)*

Two endpoints of interest:

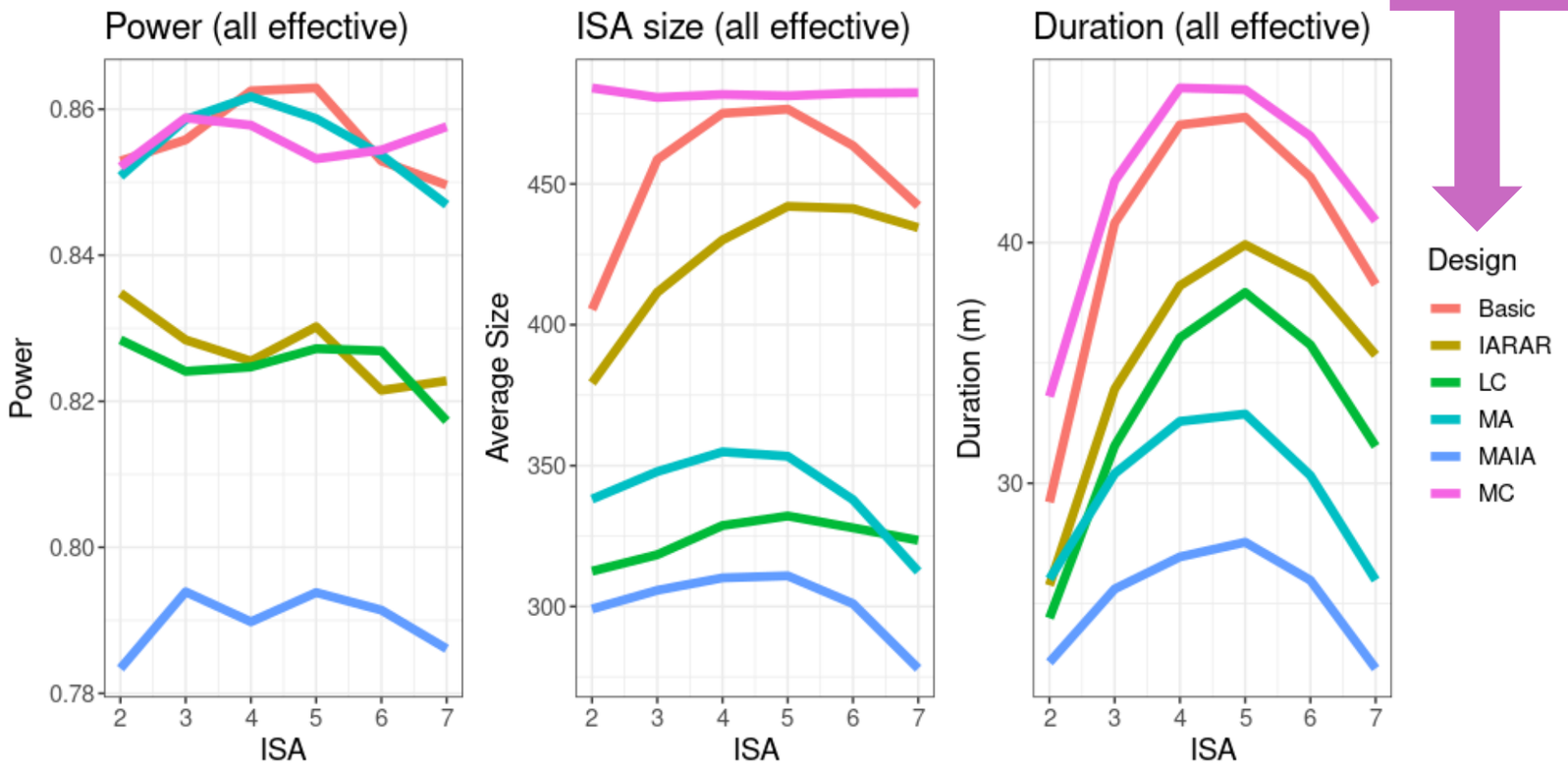
- Interim decision making based on time to culture conversion
- Final analysis: Non-inferiority vs. SoC at 12m post randomization
- Phase 2 → relaxed α of 10% one-sided



Assuming 95% favorable events: PanACEA size of 124:62 (Ctrl:Active) results in 88% power

Case 3: EU-PEARL Tuberculosis Platform

Design Comparison (All effective)



Sprinkles

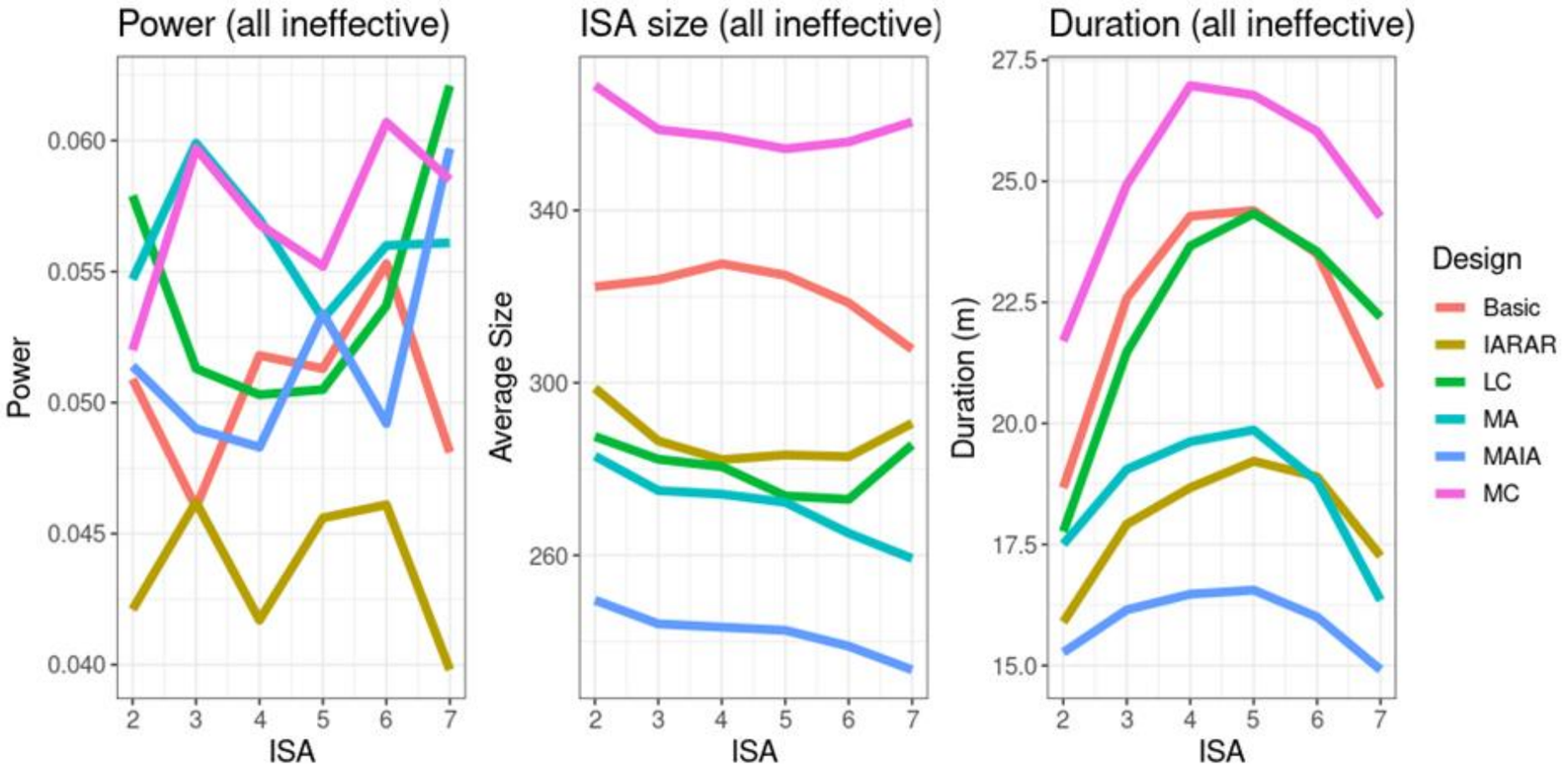
	Size	P(Ctrl)	Duration (total)
Multiple	2292	33%	*81 (23)*
Basic	1911	24%	70
MC	2001	28%	73
LC	1679	12%	63
MA	1491	29%	57
MAIA	1398	31%	55
IARAR	1845	27%	68

Which design wins?

Largest power? Smallest size? Shortest duration?

Case 3: EU-PEARL Tuberculosis Platform

Design Comparison (All ineffective)



	Size	P(Ctrl)	Duration (total)
Multiple	1902	33%	*70 (15m)*
Basic	1540	28%	58
MC	1657	35%	62
LC	1510	13%	56
MA	**1342**	**31%**	**51**
MAIA	1260	33%	49
IARAR	**1444**	**30%**	**53**

Which design wins?

Largest power? Smallest size? Shortest duration?

… under which scenario?

How useful?

Stakeholder perspectives

"How" depends on the perspective

- Patient inside trial
- Patient outside trial
- Study site
- Intervention owner (+ function)
- Regulator

Get better interventions to the patients in need fast

Blocker

?

Bottle-neck

Incentive

How to make it really useful?

Don’t start with the solution

Start with the problem identification!

- Statistical problems may not be viewed as problems by others.
- Does a platform trial address the problem?
- Which novel problems are introduced?
- Are those relevant?

What is the bottleneck?

Thank you for your attention

… and looking forward to a good discussion

… and suggestions to ensure useful platform trials


Creating a future where
disease is a thing of the past

References

Online sources:

- www.Covid19-trials.com
- www.Recoverytrial.net
- www.ispytrials.org/i-spy-platform/i-spy2
- www.stampedetrial.org/centres/information-on-stampede/

References

Publications

- Bretz, F., Schmidli, H., König, F. et al. (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. Biometrical Journal, 48, 623-634
- Krams, M., Lees, K.R., Hacke, W. et al. (2003). Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN): an adaptive dose-response study of UK-279,276 in acute ischemic stroke
- Meyer, E.L., Mesenbrink, P., Mielke, T. et al. (2021). Systematic review of available software for multi-arm multi-stage and platform trial design. Trials 22, 183
- Mok, T.S.K., Wu, Y-L., Kudaba, I. et al. (2019). Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced metastatic non-small-cell lung cancer (Keynote-042): a randomized open-label, controlled, phase 3 trial. The Lancet 393, 10183
- Park, J.W., Minetta, C.L. Douglas, Y.M. et al. (2016). Adaptive Randomization of Neratinib in Early Breast Cancer. N Engl J Med; 375:11-22
- Phillips, P.P., Dooley, K.E., Gillespie, S.H. et al. (2016). A new trial design to accelerate tuberculosis drug development: the phase iic selection trial with extended post-treatment follow-up (step). BMC Medicine, 14
- Phillips, P.P., Gillespie, S.H., Boeree, M. et al. (2012). Innovative trial designs are practical solutions for improving the treatment of tuberculosis. The Journal of infectious diseases, 205(2):250–257
- Robertson, D.S., Wason, J.M., König, F. et al. (2023). Online error rate control for platform trials. Stat Med.;
- Yao, T.J., Begg, C.B., Livingston, P.O. (1996). Optimal sample size for a series of pilot trials of new agents. Biometrics 52(3):992-1001.


janssen
PHARMACEUTICAL COMPANIES OF
Johnson&Johnson

Wrapping up…

How useful are platform trials?

Can be very useful.

… to treat, accelerate, harmonize and screen

But not always most useful.

… methods need to fit the purpose

… a standard clinical trial is the safer/cleaner approach

… but possibly less efficient.