Not all Platform Trials are Created Equal:

lessons from clinical trials in outbreak settings

Lori E Dodd, PhD
Clinical Trials Research Section
Biostatistics Research Branch
NIAID
Principal statistician for the PALM1 Ebola RCT and the ACTT COVID-19 trials

Table 1. Types of Master Protocols.

Type of Trial	Objective
Umbrella	To study multiple targeted therapies in the context of a single disease
Basket	To study a single targeted therapy in the context of multiple diseases or disease subtypes
Platform	To study multiple targeted therapies in the context of a single disease in a perpetual manner, with therapies allowed to enter or leave the platform on the basis of a decision algorithm

Woodcock and LaVange, NEJM, 2017

Platform Trials

"Answer more questions, more efficiently and in less time†"

Shared resources allow for

Common control group

- Sample size efficiency

Larger study -> greater statistical power

-> more data on subgroups

Efficient use of study related procedures:

- Screening & enrollment procedures
- Sites
- Systems
- Data and safety monitoring board
- Master protocol with appendices for added arms

† Woodcock and LaVange, NEJM, 2017

Platform Trial Decisions

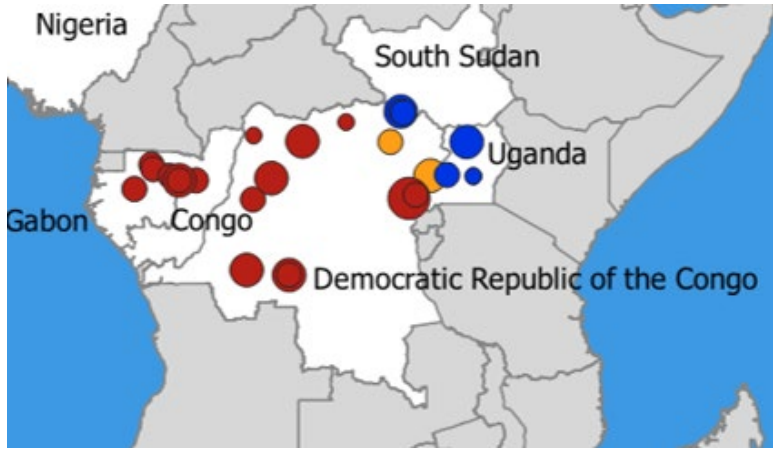
- Number of arms to include
- Adjustments for multiple comparisons
- When and how to add arms?
- When and how to drop arms?
- Defining the control group
- Trade-off between quantity and quality

Number of study agents/arms

- Number of candidate interventions
  - Candidates may differ in terms of pre-existing evidence of potential efficacy
- More study arms means fewer participants in any given arm
- Number of expected participants, often unknown
  - Target effect size, study power and type I error rate considerations drive sample size
  - If number of participants could be known from the outset, one could optimize study design
- Duration of study
  - Trade-off between obtaining an answer to one questions sooner versus many answers later

Example: PALM1 Ebola RCT in the DRC

10th DRC outbreak

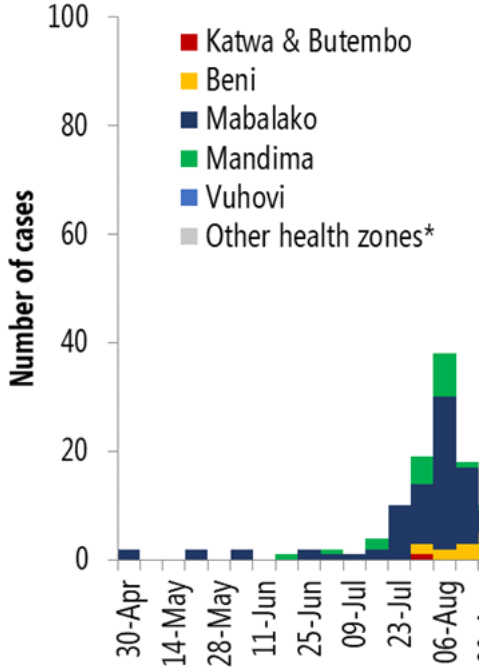


Aug 14, 2018: protocol concept

~80 cases reported

Previous outbreaks ended with an average of 133 cases (range 1 – 318)

Potential enrollment numbers uncertain



Candidate treatments:

ZMapp – monoclonal antibody studied in West Africa outbreak (control)

mAb114 – new monoclonal antibody

REGN-EB3 – new monoclonal antibody

Remdesivir – antiviral

How many arms can we study?

Example: PALM1 Ebola RCT in the DRC

Primary Endpoint

28-day mortality

Sample Size

125 per arm for 85% power using a 2-sided type I error rate= 0.05.

30% mortality in control vs 15% experimental arm.

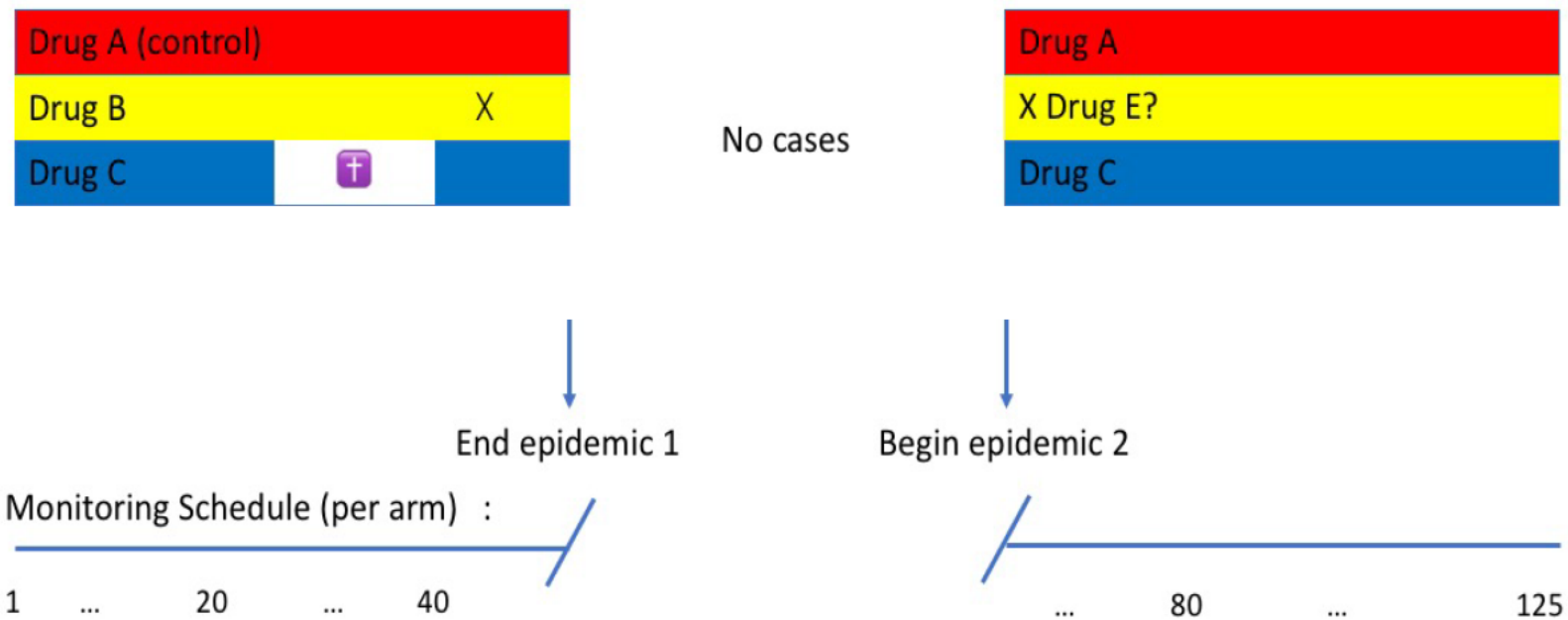
How many arms?

250 for 2 arms

375 for 3 arms

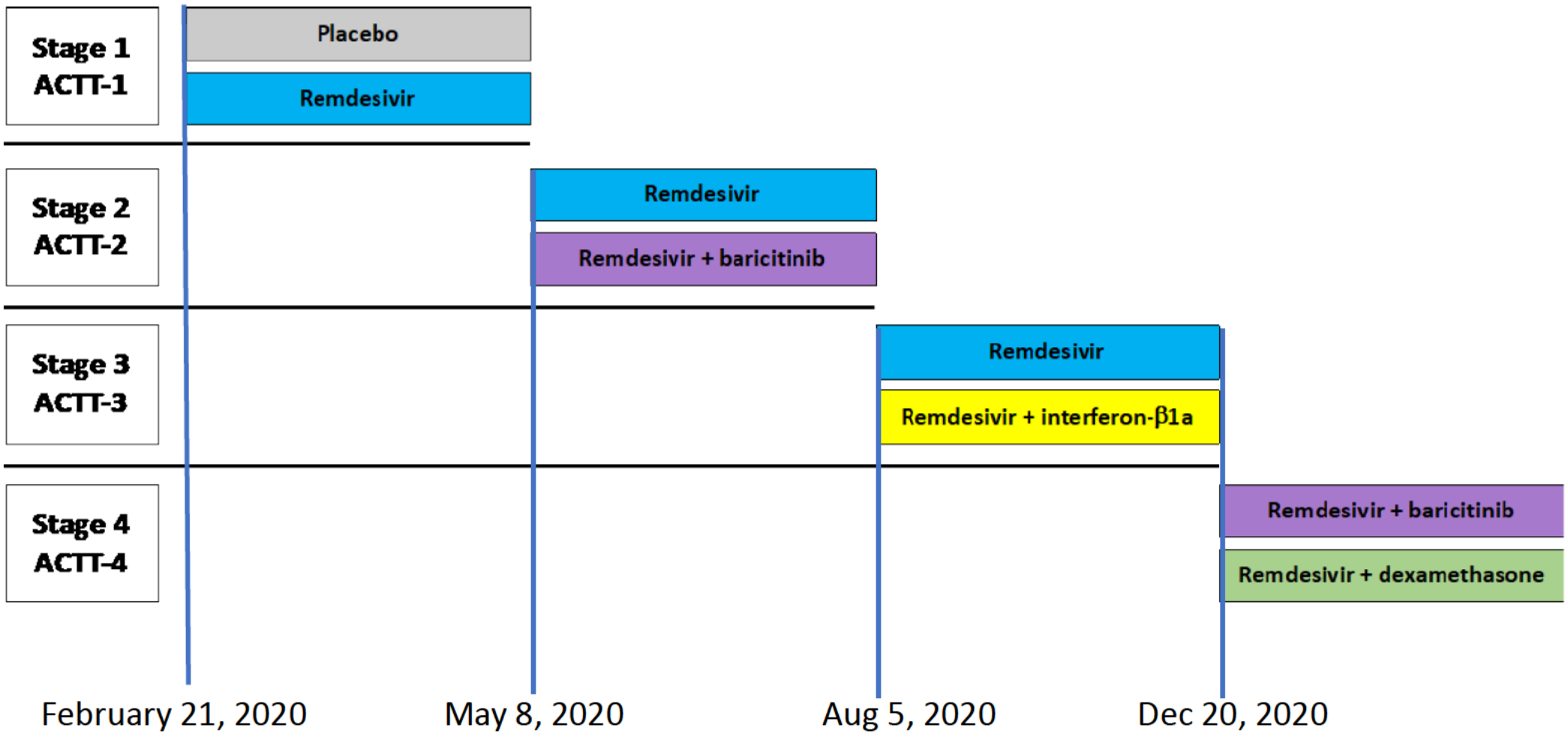
500 for 4 arms

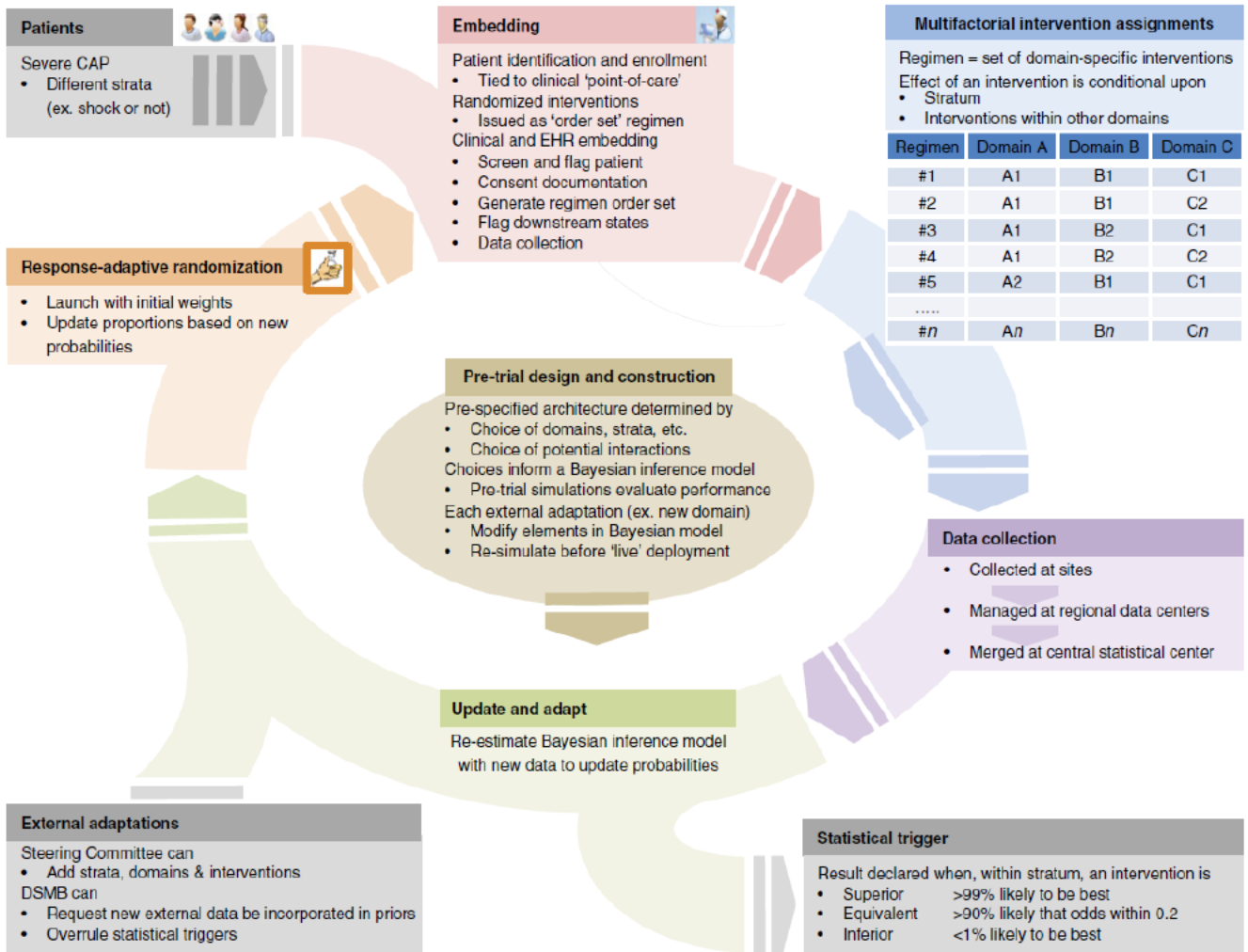
Sample size advantage: 2 separate trials -> sample size 500



Example: Adaptive COVID-19 Treatment Trial (ACTT)

Early in the pandemic remdesivir was the only intervention with known activity against SARS-CoV2
Urgency of situation meant answer about the most promising agent needed quickly





Example: REMAP-CAP

Type I Error Rate Control

Multiplicity issues must be addressed

Multiple hypothesis tests will inflate the overall Type I error rate

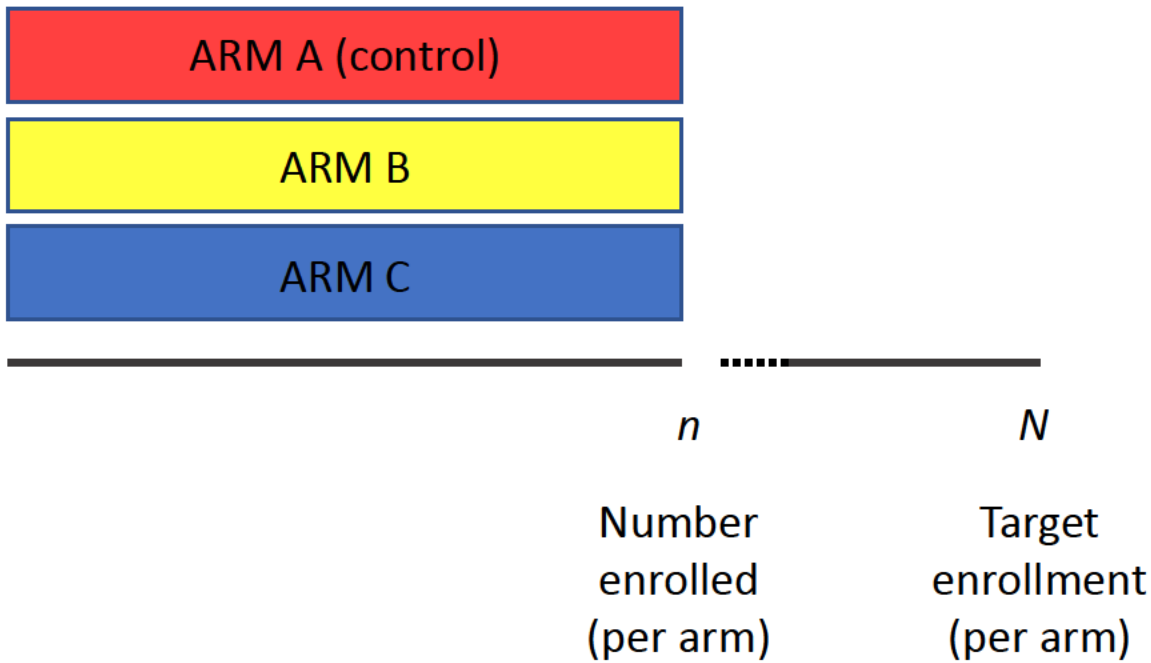
- You get “multiple bites at the apple”
- Different drugs from different companies
  - Valuable to have different drugs evaluated in the same study
  - Strict multiplicity adjustments may disincentivize company participation
- Some argue multiple comparison adjustments are not necessary but it depends.
- Same drug but multiple arms are different doses => more strict adjustment
- Dunnett’s test for comparisons to a common control => too strict?
  - Platform trial power advantage remains: “When the total number of subjects is the same in a single versus separate experiments, power is generally higher in a single experiment even if a Dunnett adjustment is made,” Follmann and Proschan, 1994
    - Recommendation to enroll more controls than exp’l arms
    - Strategy will reduce power for comparisons of exp’l arms, which is often of interest

PALM1 multiple comparison adjustments:

- Product developers have little incentive to participate in a platform trial that adjusts for multiple comparisons
- With 2 primary comparisons: Dunnett's pretty extreme/similar to a Bonferroni adjustment
- In Ebola setting, achievable sample size is unpredictable and limited
- Increase in type I error rate beyond traditional/arbitrary 0.05 acceptable in this setting – examples in rare cancers apply
- Consider each comparison as a separate study
- No adjustment for multiple comparisons planned for PALM1
- US FDA concurred with this approach for this setting
- Not applicable to all settings.
  - Each setting must be considered separately

Adding Arms to a PLATFORM trial

Adding arms to an ongoing trial



Consider adding a new arm *D*

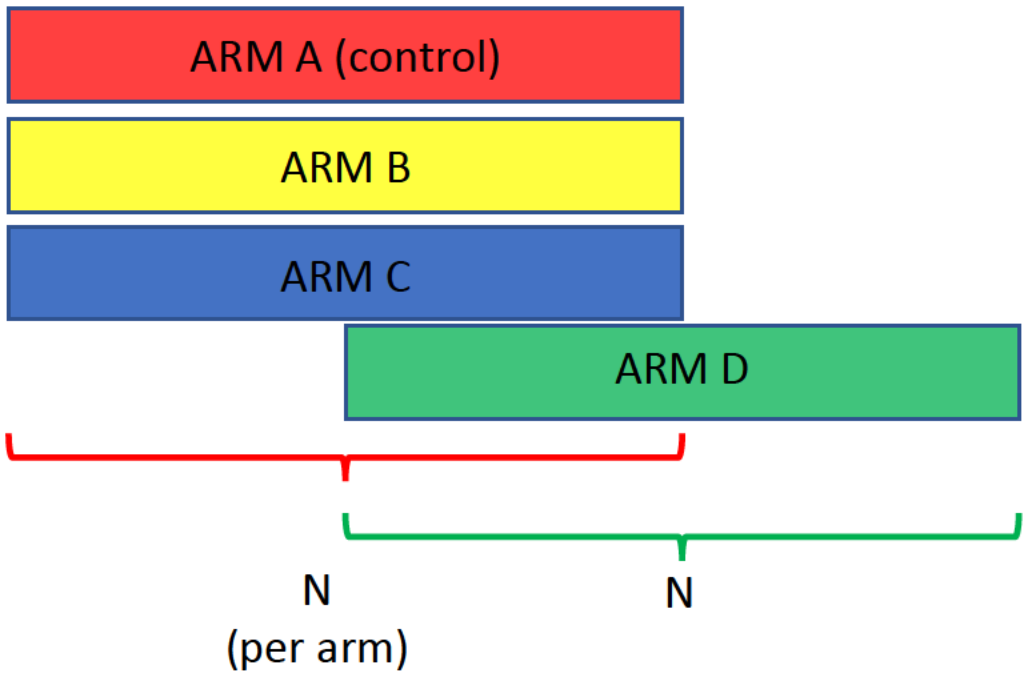
- *n* already enrolled in control arm
- *N* total sample size (per arm) for desired type I/II error rates

Adding arms to an ongoing trial

How will addition of a study arm influence the existing arm comparisons?

- For a fixed total sample size, power will be reduced
- Delayed conclusion about efficacy of Arm *B* and Arm *C*
- Power for the new arm will lag relative to those already under study

Randomization strategies when adding arms



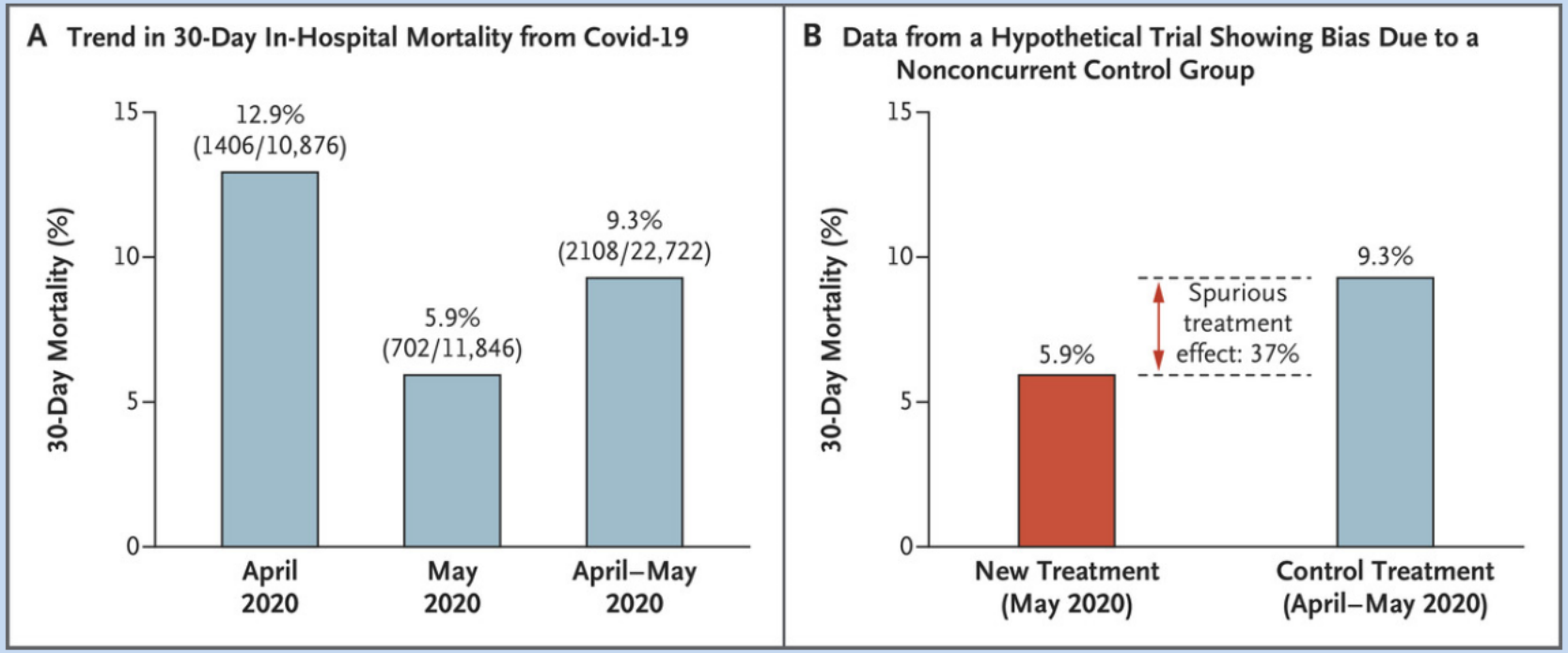
Increase sample size by N for Arm D

Randomize 1:1:1:1

Include all control participants in comparison of new arm

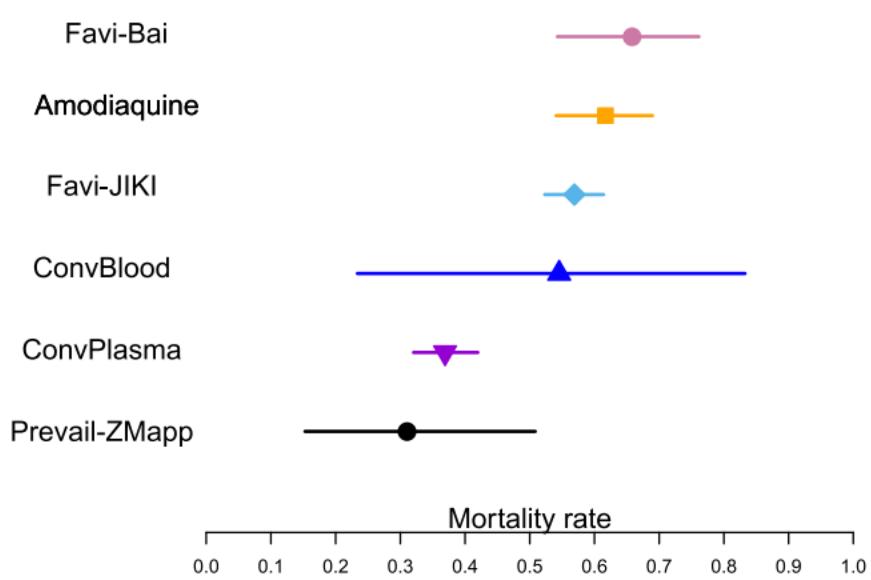
- Total sample size increases by N
- Analogous to using *n* observations from historical control data in the comparison of *D:A*
- Time trends may bias comparison of added arm
- What to do with other study arms if one arm proves superior?

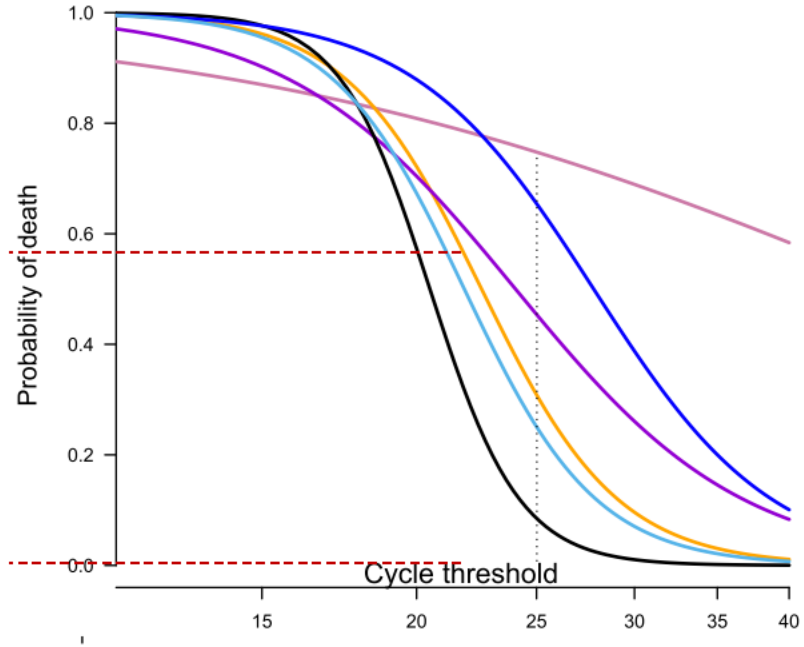
Hypothetical Example of How Nonconcurrent Randomization Could Bias the Results of a Trial.

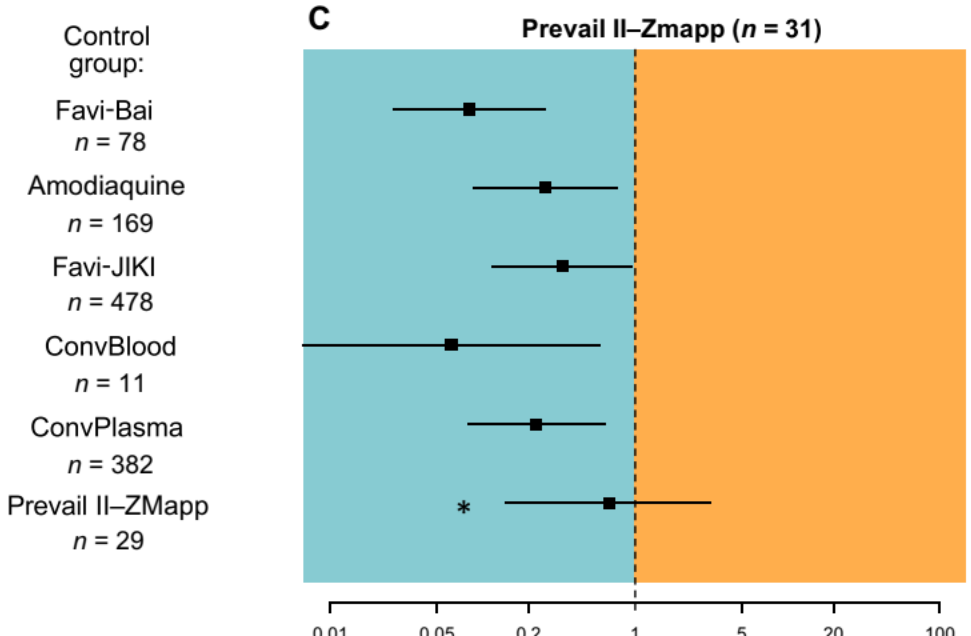


LE Dodd et al. N Engl J Med 2021;384:1572-1573.

Meta-analysis from West African Ebola studies







Dodd et al, *Sci Trans Med, 2019*

Possible explanatory variables:

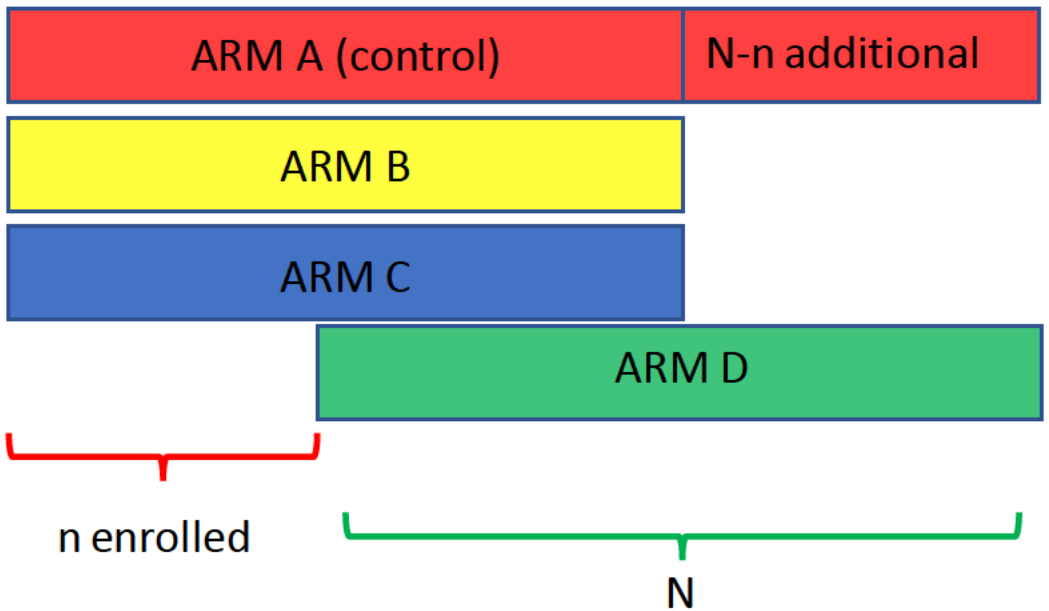
- Differences in supportive care
- Different assay platforms
- Treatment facility differences
- Viral load at time infection
- Time from onset to enrollment
- Co-morbidities
- Host genetics
- Age
- Sex
- Changes over time of the above variables
- Viral evolution

Importance of concurrent controls

- Outbreak diseases may have considerable heterogeneity in natural course of disease over time
- Standard of care changes rapidly and may dramatically alter “usual care” outcomes (e.g. Ebola outbreak in 2014-2016)†
- Study arms need equivalent patient populations (on average) so that differences between arms can be attributed to exp’l treatment
- Validity relies on concurrent controls– differences between arms are approximately the same over time (for population)
- Designs that do not keep an equal allocation of controls over time may lead to errant conclusions

†Dodd et al, *Sci Trans Med, 2019,* DOI: 10.1126/scitranslmed.aaw1049

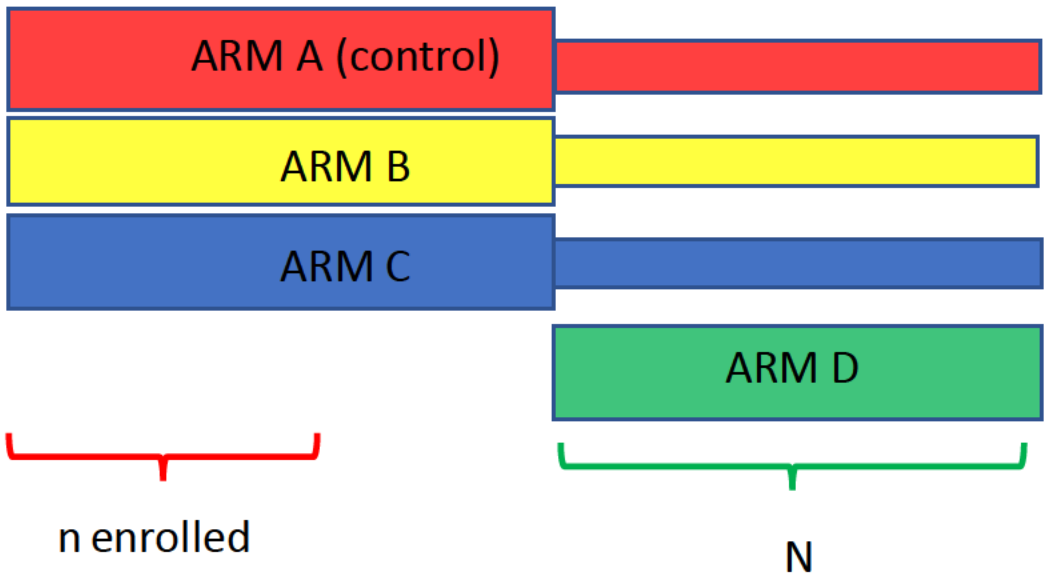
Randomization strategies when adding arms



- Analyses include only those concurrently randomized
- Total sample size increases by N+ n

- May not want to stop enrollments in B & C
- What to do with other study arms if one arm proves superior?

Randomization strategies when adding arms



Modification

Update randomization probabilities at the time the arm is added so that all four arms complete enrollment at the same time.

Potential time trend bias.

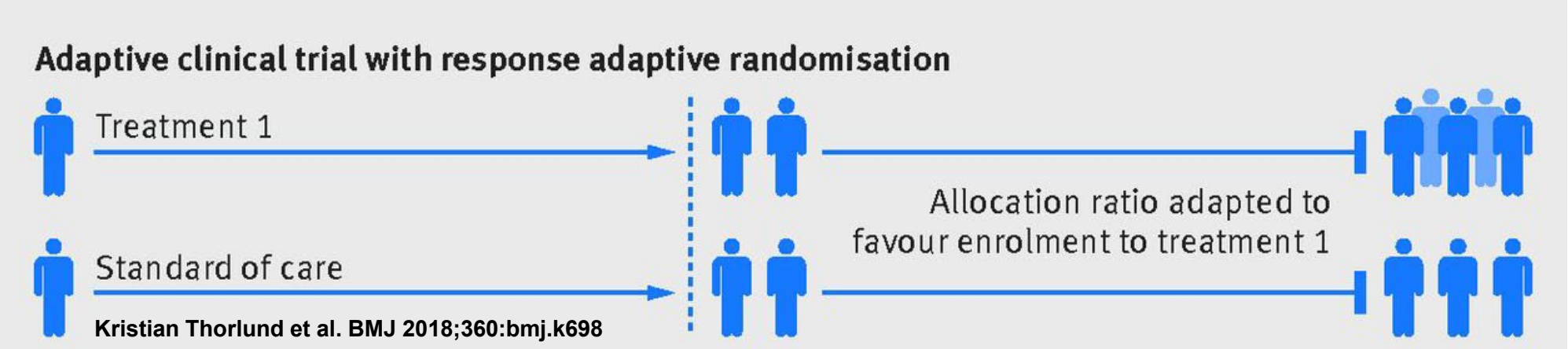
Stratified analyses for proper inference for Arm B & C comparisons.

- -Models of time-trend proposed but validity depends on the ability to accurately model time-effect.

What to do with other study arms if one arm proves superior?

Randomization strategies when adding arms

Response adaptive randomization (RAR)



- With RAR, randomization probabilities are updated based on outcomes, typically restricted to “burn-in” periods set by a certain number enrolled.
- Adding an arm: “burn in” period designated for new arm
- Existing data and pre-specified functions guide updates to randomization probabilities
  - Various approaches exist: Ventz et al (Biostatistics, 2018), Berry et al (Clin Trials, 2016)
- Concern over time trends bias

Clinical Infectious Diseases

INVITED ARTICLE

IDSA Infectious Diseases Society of America

hivma hiv medicine association

OXFORD

INNOVATIONS IN DESIGN, EDUCATION AND ANALYSIS (IDEA): Victor De Gruttola and Scott R. Evans, Section Editors

Resist the Temptation of Response-Adaptive Randomization

Michael Proschan[1] and Scott Evans[2]

Response-adaptive randomization (RAR) has recently gained popularity in clinical trials. The intent is noble: minimize the number of participants randomized to inferior treatments and increase the amount of information about better treatments. Unfortunately, RAR causes many problems, including (1) bias from temporal trends, (2) inefficiency in treatment effect estimation, (3) volatility in sample-size distributions that can cause a nontrivial proportion of trials to assign more patients to an inferior arm, (4) difficulty of validly analyzing results, and (5) the potential for selection bias and other issues inherent to being unblinded to ongoing results. The problems of RAR are most acute in the very setting for which RAR has been proposed, namely long-duration "platform" trials and infectious disease settings where temporal trends are ubiquitous. Response-adaptive randomization can eliminate the benefits that randomization, the most powerful tool in clinical trials, provides. Use of RAR is discouraged.

Another good reference: Korn, EL, Freidlin, B. Outcome-adaptive randomization: is it useful? J Clin Oncol 2011; 29: 771–776

Perspective

Time trends with response-adaptive randomization: The inevitability of inefficiency

CLINICAL TRIALS

Korn and Freidlin, 2022

Inflated type I error rate with change of randomization allocation ratio and time trends in outcomes

Table 2. Simulated levels and powers[a] of three analysis methods for detecting a treatment difference for a randomized trial with 200 patients in each of two time blocks (nominal one-sided type I error = 0.05, standard normally distributed errors).

			Analysis method			
Treatment effect	Effect of second time block	Randomization ratio for second time block[b]	Z-test[c] (unstratified)	Time-block stratified Z-test[d]	Re-randomization test (unstratified test statistic)[e]	Re-randomization test (stratified test statistic)[f]
0	0	1:1	0.049	0.049	0.049	0.049
	0	9:1	0.050	0.050	0.049	0.049
	0.30	1:1	0.049	0.049	0.049	0.049
	0.30	9:1	0.365	0.050	0.049	0.049

CLINICAL TRIALS

Time trends with response-adaptive randomization: The inevitability of inefficiency

Korn and Freidlin, 2022

Table 2. Simulated levels and powers[a] of three analysis methods for detecting a treatment difference for a randomized trial with 200 patients in each of two time blocks (nominal one-sided type I error = 0.05, standard normally distributed errors).

			Analysis method			
Treatment effect	Effect of second time block	Randomization ratio for second time block[b]	Z-test[c] (unstratified)	Time-block stratified Z-test[d]	Re-randomization test (unstratified test statistic)[e]	Re-randomization test (stratified test statistic)[f]
0.27	0	1:1	0.853	0.852	0.850	0.849
	0	9:1	0.795	0.716	0.714	0.713
	0	7:3→7:3[g]	0.796	0.795	0.793	0.791
	0.30	1:1	0.845	0.852	0.843	0.849
	0.30	9:1	NA[h]	0.716	0.707	0.713
	1.00	1:1	0.776	0.852	0.773	0.849
	1.00	9:1	NA[h]	0.716	0.646	0.713

Example: Adding an arm in PALM1

On ***November 20, 2018*** enrollment started

- Increasing pressure to add REGN-EB3, a triple monoclonal antibody treatment.

On ***January 26, 2019*** REGN-EB3 was added

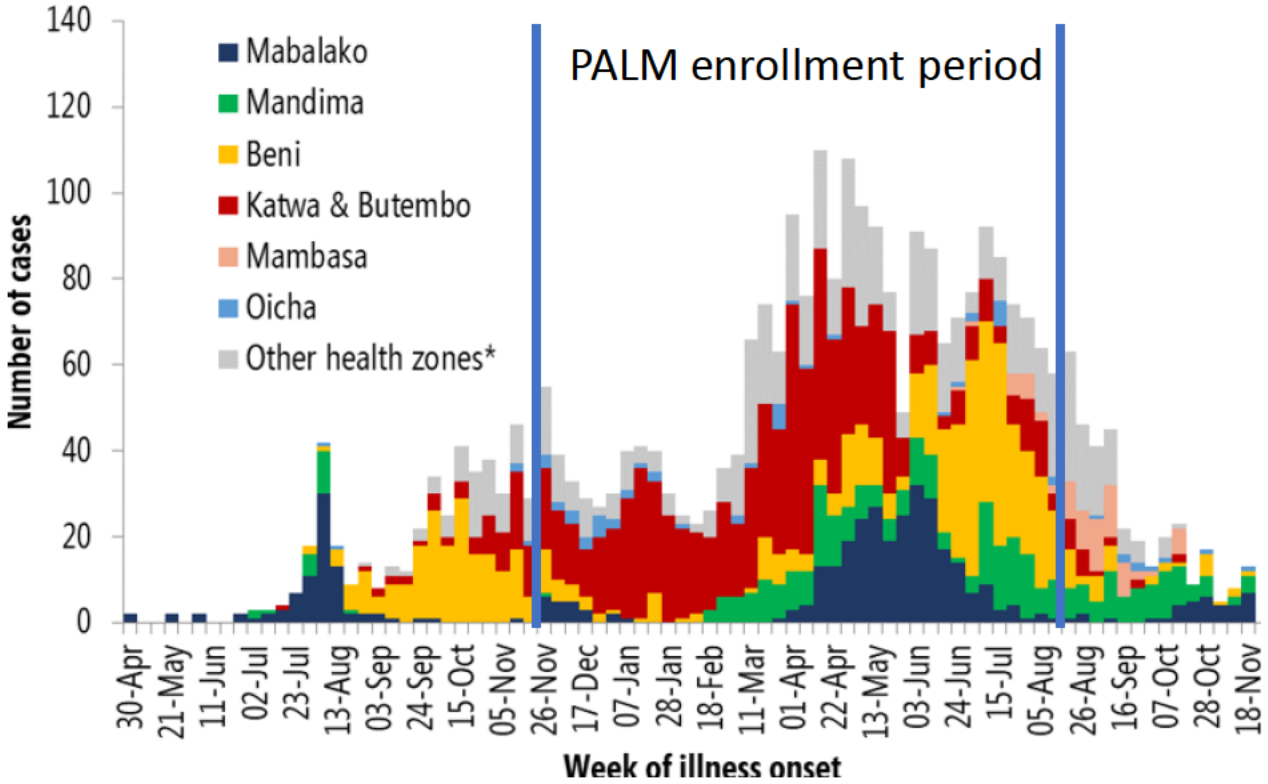
- At the time there were 51 enrolled, 15 of them were on the control arm
- 15 fewer controls for the REGN-EB3 to control comparison not considered enough to raise concerns about power
- Randomization allocation set to 1:1:1:1
- What would we have done if there had been 30 controls? 50 controls?

Dropping Arms from a Platform Trial

When to drop arms?

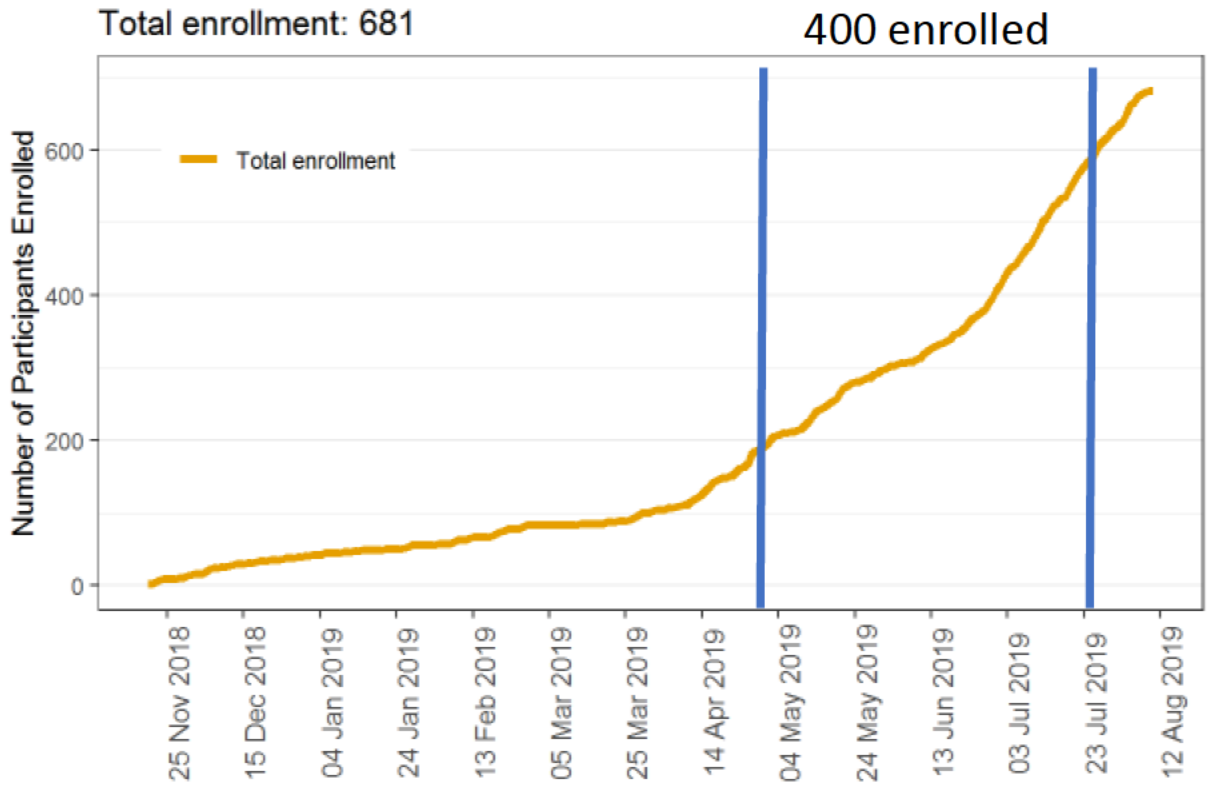
What if Arm *B* crosses an efficacy boundary but Arm *C* and *D* have not?

- Stopping the control arm will prevent accumulation of additional information about the efficacy of Arms *C & D*.
- Is it ethical to continue randomization to the control arm?
- Does it matter if the Arm *D* is close to crossing an efficacy boundary?
- Should the recommendation depend on the importance of the medical community having more than one treatment option?



PALM1: Ebola treatment RCT

Time from enrollment until 28-day mortality results are in database: 5-6 weeks.



PALM1: Interim Monitoring

- Mid-July: learn of rumors that one therapeutic candidate is harmful
  - Increased concern amongst field team
- August 9th meeting: 681 enrolled
- Only 376 patients with a day 28 outcome in time for DSMB report
  - ~50% of total information, enrollment 94% complete
- Study team, including site staff, well aware of this disparity and concerned. The study statistician was concerned too!
- Looking at earlier timepoint (e.g., day 10) not considered in protocol

What we (the study team) didn’t know

- Fewer than 5% of deaths occurred after day 10.
- Early on the DSMB requested analysis on a “surrogate” endpoint
  - For participants who have not reached day 28 but have been on study at least 10 days, include their most current status.
  - Compare mortality proportions based on this surrogate

PALM1 Aug 7, 2019 DSMB meeting

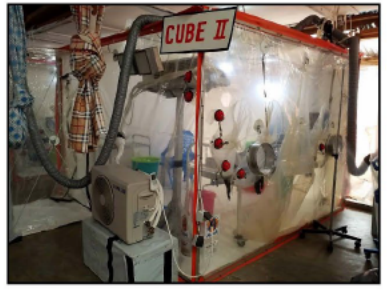
- Using 10-day mortality, a near-perfect surrogate for 28-day mortality, interim results showed that the two leading study arms (mAb114 and REGN-EB3) would be statistically significant after the final 28-day outcome data were available.
- The DSMB recommended stopping random assignment to the control arm and remdesivir and continuing follow-up to 28 days for all patients randomly assigned by August 7, 2019
- The sponsor (the National Institute of Allergy and Infectious Diseases) accepted the recommendation and transitioned new patients to the proven treatments.

Mulangu et al, NEJM 2019
Dodd & Proschan, NEJM Evidence 2022

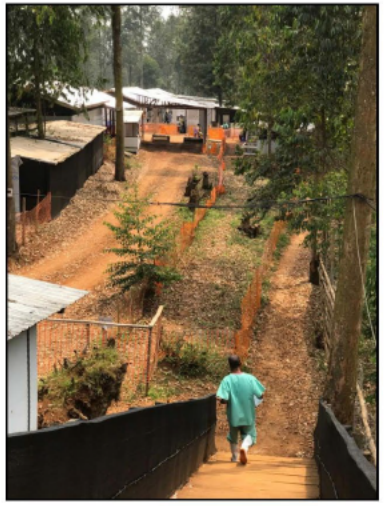
DRC-US Joint Clinical Research Partnership

Established November 2018

Katwa



Butembo



PALM Clinical Trial Sites

Mangina

Beni

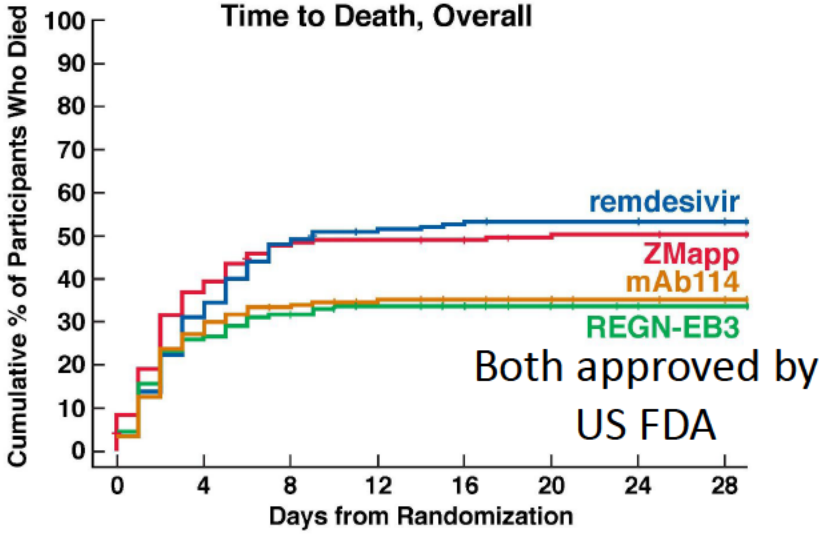
The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

December 12, 2019

Randomized, Controlled Trial of Ebola Virus Disease Therapeutics

PALM Consortium Study Team



Both approved by US FDA

December 20, 2019

Science

SPECIAL ISSUE

sciencemag.org

AAAS

Runner-up for the 2019 Breakthrough of the Year

PALM Trial: Hope for Ebola Patients, At Last

"In the midst of another outbreak, the deadliest in the DRC's history, scientists finally identified two drugs that dramatically reduced death rates from the disease.... Simply conducting the trial was a notable achievement in itself."

David Sackett 2019 Trial of the Year Award

Is there value in having more than one trial ?

- Yes
- Additional trial(s) add external validity
- Regulatory requirement for two adequate and well-controlled clinical investigations (see FDA guidance)†
- Data in other settings/populations of value
- Additional data may contribute to knowledge about subgroups (via meta-analysis)
- Value depends on the quality of the other trials and the extent to which they are harmonized

† https://www.fda.gov/media/133660/download

	ACTT-1	SOLIDARITY
	Placebo	Standard care
	Remdesivir	Remdesivir
Design	Placebo, double-blind, adaptive, randomized controlled	Randomized controlled not blinded; large/simple trial
Inclusion Exclusion	Hospitalized RT-PCR+ (≤ 72 hrs) Lower resp. disease Any O_2 level	18+ years old Admitted to hospital with definite COVID-19
Randomization	1:1, stratified severity & site	1:1
1º Endpoint	Time to recovery (28-days)	In-hospital mortality
Countries	11 countries	35 countries
Enrollment dates	February 21- April 19, 2020	March 22, 2020-Jan 29,2021
Sample size	1062	1st:2743 (RDV) + 2708 (C) Update: 4169 (RDV) + 4151 (C)
Conclusions	Improvement in time to recovery (med 5 days); most prominent in low-O2 group, with suggested mortality benefit; improvements in progression to ventilation or death *regulatory approvals	NEJM 2020: No benefit in mortality; time-to-discharge biased because trial not blinded. Lancet 2021 update: improvements in mortality and ventilation or death in low O2 group

Confusion over treatment guidelines

Differences in recordation of respiratory support:
ACTT records high-flow O_2
Solidarity does not

In 2020:

NIH Treatment Guidelines recommends remdesivir use in patients requiring low-oxygen supplements patients

WHO Treatment Guidelines recommends not using remdesivir. Guidelines updated later to recommend remdesivir use

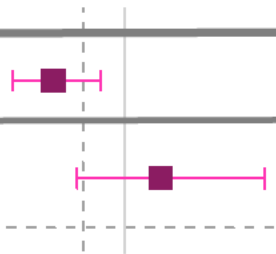
Larger/Simple vs Smaller/Complex Trials

"Large and simple" does not necessarily mean better

	Larger & simple Example: Solidarity		Smaller & Complex Example: ACTT-1
+	Large sample sizes/more power	-	Smaller studies/less power
+	Less burden on site staff per participant	-	More burden on site staff and study team/participant
-	Less likely to be placebo-controlled	+	More likely to be placebo-controlled
-	Primary endpoint: in-hospital mortality	-	Primary endpoint: time to discharge
-	Less likely to be sufficient for FDA approvals	+	More likely to be sufficient for FDA approvals
-	Little/less safety data	+	More safety data; ability to assess signals of harm, stop arms, protect participants
-	Less data on secondary endpoints	+	More data on secondary endpoints***: evidence about treatment efficacy comes from much more than the primary endpoint***
-	Less granular data; limited data targeted on primary endpoint. Data collection at fewer time points	+	More granular data can inform other analyses, studies, exploration of mechanism of action
-	Size limits ability quality oversight Less data quality checking/monitoring overall	+	More data quality checking/monitoring; higher quality studies

Forest plot presenting meta-analysis analyses for the mortality

Subgroup	Total number	Remdesivir	No Remdesivir	aOR	ll	ul	p–int
No O_2/Low O_2	8059	404/4094	461/3965	0.81	0.702	0.935	0.022
High-flow, Mech Vent, ECMO	1685	253/841	241/844	1.10	0.873	1.373	



Abbreviations: aOR, adjusted Odds Ratio; ECMO, extracorporeal membrane oxygenation; ll, lower limit; p-int, p-interaction; ul, upper limit

**p*-value interaction from the main model using the covariate as continuous linear interaction term

Effects of Remdesivir in Hospitalized Patients with COVID-19: Systematic Review and Individual Patient Data Meta-Analysis of Randomized Clinical Trials.
Amstutz et al.
Lancet preprint server: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4244759

Conclusions

- COVID-19 has provided an unprecedented number of clinical trials and participants over a relatively short duration.
- Trials should seek to balance the advantages of a large/simple trial with the benefits of smaller/complex trials.
- Response adaptive randomization may lead to inflated type I error and statistical inefficiency-> avoid for definitive clinical trials
- Definitive results about any arm will effect ability to collect data on other arms.
- Adding an arm during a study presents challenges
- Trade-off in statistical power with many arms
  - Answer one question faster?
- Multiplicity adjustments need careful consideration for each setting