Good Morning Meeting

OPTIMIZING THE TRANSITION FROM EARLY TO LATE PHASE TRIALS IN ONCOLOGY

WELCOME!

Scan the QR code to view program and participants

We will begin at 9:00

IDDI
PASSION. SCIENCE. EXPERIENCE.

Medicon Valley Alliance

IDDI

PASSION. SCIENCE. EXPERIENCE.

OPTIMIZING THE TRANSITION FROM EARLY- TO LATE-PHASE TRIALS IN ONCOLOGY

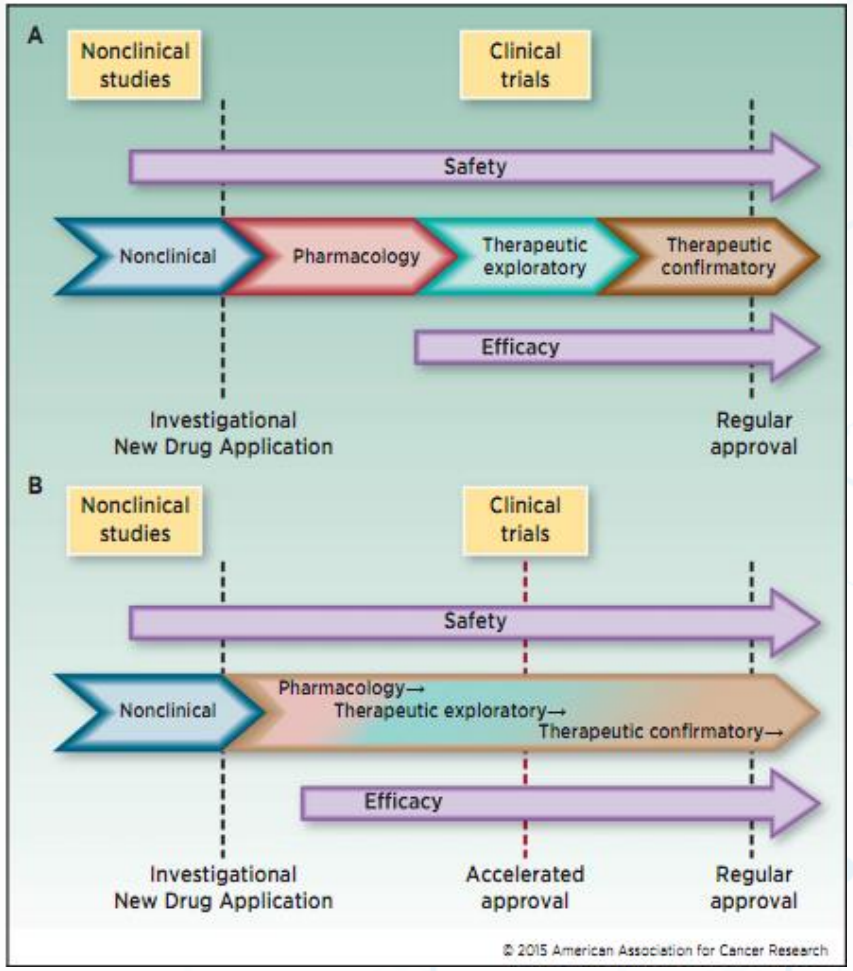
EVERARDO D. SAAD, MD
MEDICAL DIRECTOR

OUTLINE

- Some definitions
- Issue #1: regression toward the mean
- Issue #2: focus on subgroups
- Issue #3: misinterpretation of *P*-values
- Issue #4: reliance on historical data
- Issue #5: targeting accelerated approval
- Conclusions

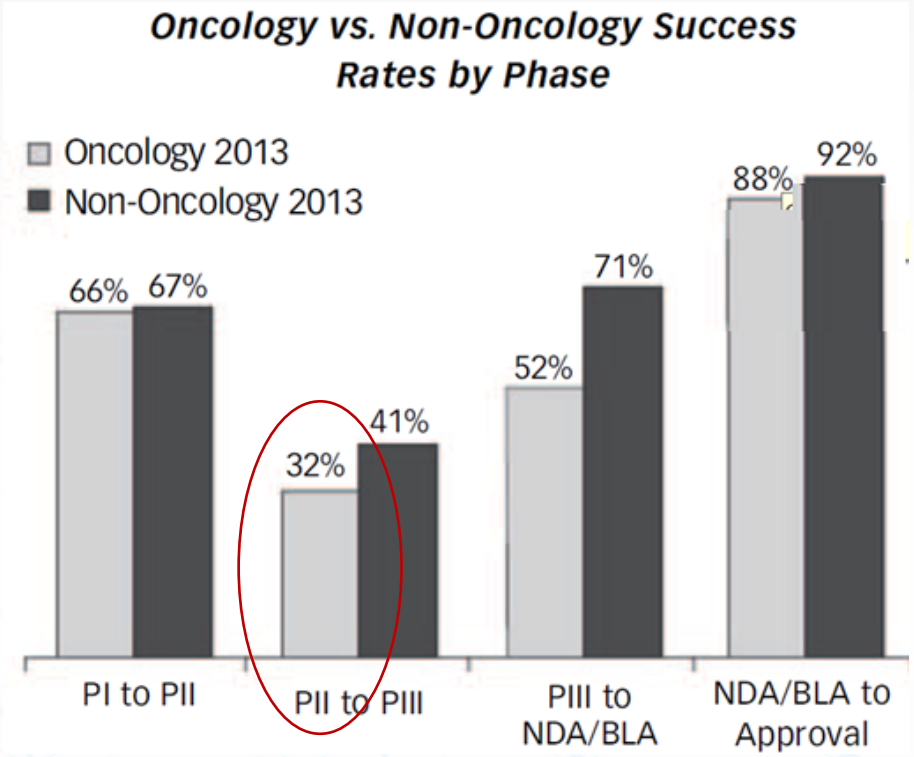
EARLY AND LATE PHASES OF DEVELOPMENT

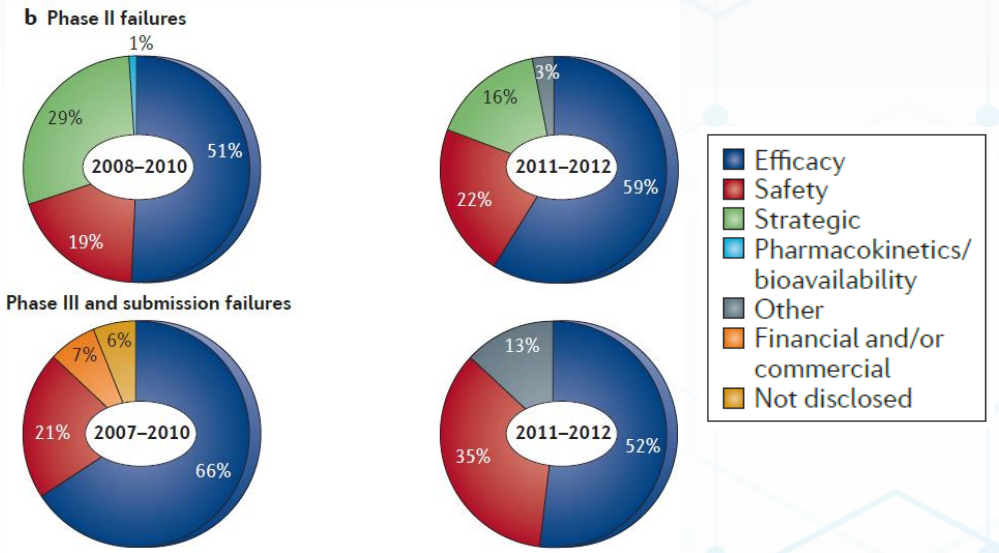
- Early
  - Phase 1 trials, whether or not first-in-human
  - Phase 1 trials with expansion cohorts, phase 1/2 trials
  - Single-arm phase 2 trials
  - Randomized phase 2 trials with no comparative intent
- Late
  - Randomized phase 2 trials with comparative intent
  - Phase 2/3 trials
  - Phase 3 trials



Theoret et al, Clin Cancer Res 2015;21:4545-51.

FAILURE IN THE TRANSITION





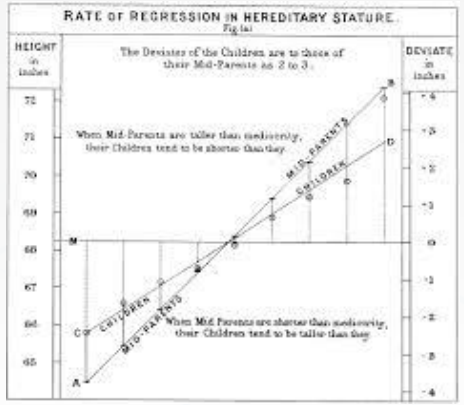
Grignolo et al, Applied Clinical Trials 2016; 25(8):36-42;
Arrowsmith et al, Nat Rev Drug Discov 2013;12:569.

TYPICAL CULPRTIS

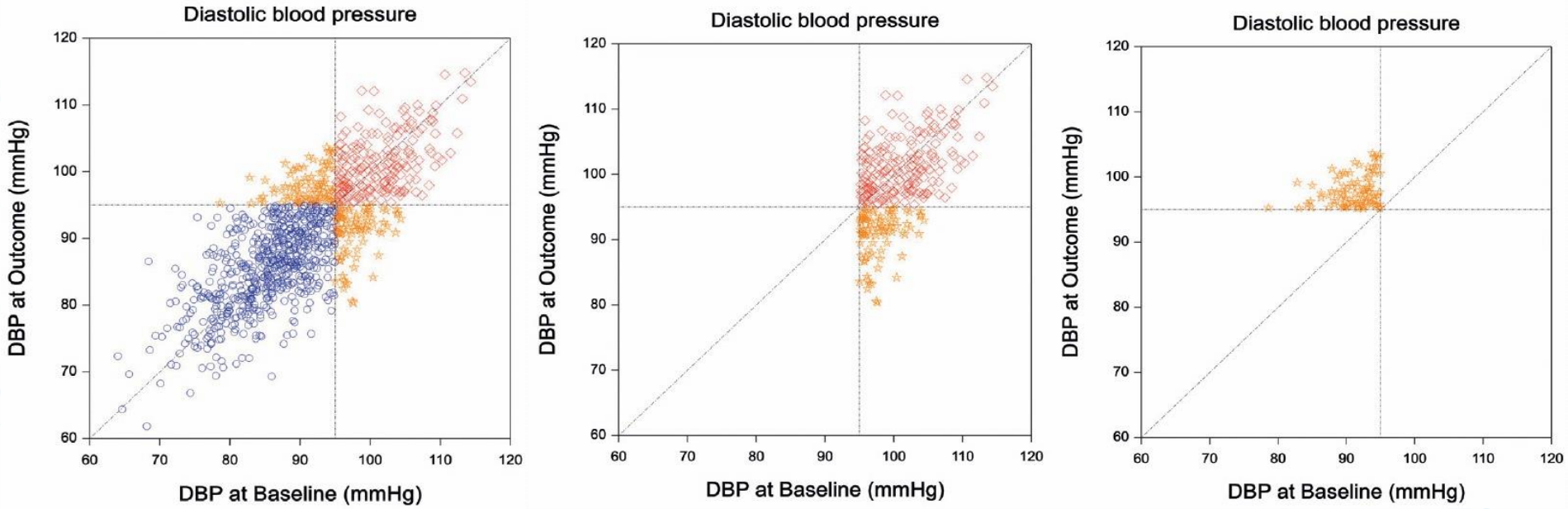
DRIVERS OF FAILURE	EXAMPLES
Inadequate basic science	• Beneficial effects in animal models not reproduced in humans • Poor understanding of target disease biology
Flawed study design	• Patient population definition changed from Phase II to Phase III • Phase II surrogate endpoint not confirmed by Phase III clinical outcomes • Insufficient sample size
Suboptimal dose selection	• Inadequate dose finding in Phase II • Poor therapeutic indices
Flawed data collection and analysis	• Phase II "false positive" effects were not replicated in Phase III • Overoptimistic assumptions on variability and treatment difference • Missing data; attrition bias; rater bias • Wrong statistical tests; other statistical issues
Problems with study operations	• Data integrity issues; GCP violations • Recruitment, dropouts, noncompliance with protocol • Missing data; unintentional unblinding
Other	• Insufficient landscape assessment of current standard of care and precedents

Grignolo et al, Applied Clinical Trials 2016; 25(8):36-42.

ISSUE #1: REGRESSION TOWARD THE MEAN



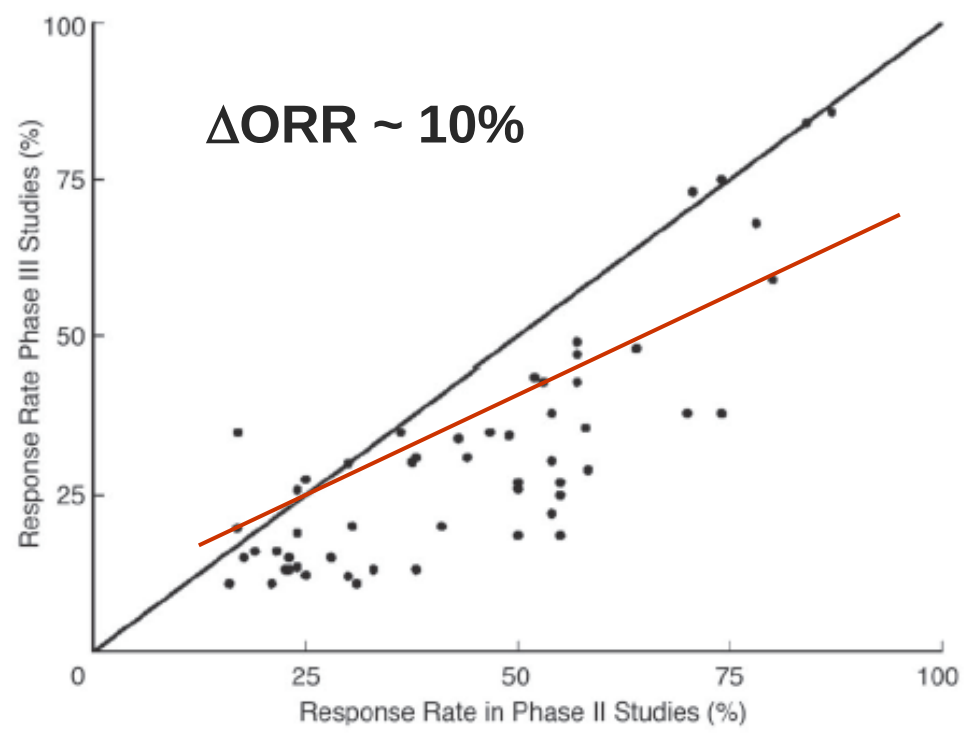
When mid-parents are taller than average, their children tend to be shorter than they; when mid-parents are shorter than average, their children tend to be taller than they.

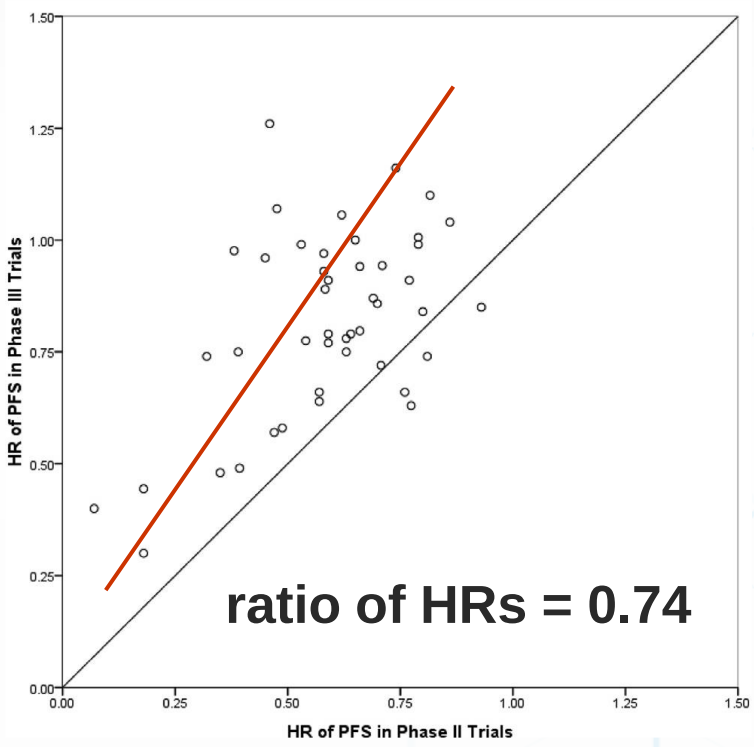


Galton, J Anthropological Institute of Great Britain and Ireland 1886;15:246-63; Senn, Significance, September 2011;126.

ISSUE #1 IN DRUG DEVELOPMENT

- Only “positive” phase 2 trials lead to phase 3 trials
- But treatment effects in phase 3 trials are expected to be smaller, on average, than in the preceding phase 2 trials

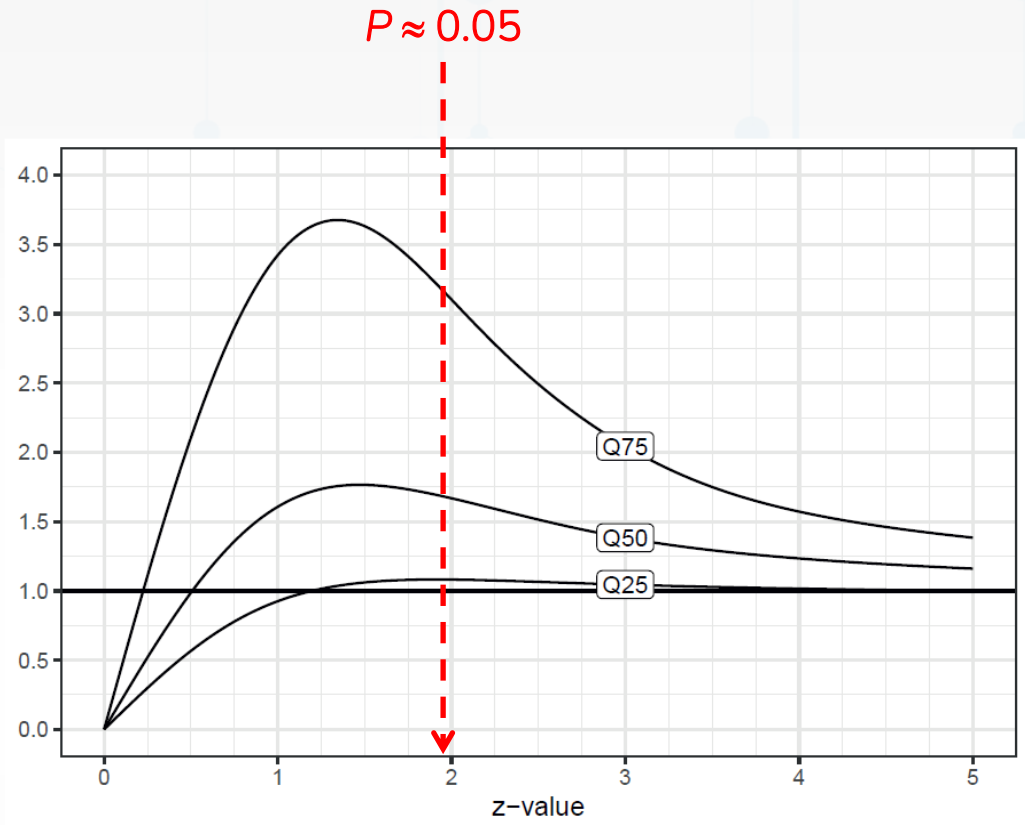




Zia et al, J Clin Oncol 2005;23:6982-91; Liang et al, Eur J Cancer 2019;121:19-28.

ISSUE #1: WHAT TO DO?

- Acknowledge the fact
- Consider the *P*-value (see Issue #3)
- "Discount" results from early phases
  - Informally
  - Formally
    - For example, correcting for the "exaggeration ratio" (R), the ratio of the observed to the true treatment effect
    - R can be estimated using the observed treatment effect and its standard error (their ratio is a z-value) and a database of trial results, e.g., the Cochrane Database of Systematic Reviews or a more targeted database if available



van Zwet et al, Significance, December 2021;16-21; Kirby et al, Pharm Stat 2012;11:373e85; https://vanzwet.shinyapps.io/shrinkrct/

ISSUE #2: FOCUS ON SUBGROUPS

- Subgroup analyses are potentially misleading, but inevitable
- They are plagued by “increased” type I and type II errors
- Even interaction tests can mislead, in absence of strong biologic plausibility
- Don’t forget, one might conclude from ISIS-2 that aspirin doesn’t work in libras and geminis!

With k independent subgroups and no difference in treatments, the probability of at least one significant subgroup is $1-(1-\alpha)^k$

Thus, if $\alpha=0.05$, $k=5$, $\text{Prob}=1-(1-0.05)^5=0.27$

Lancet 1988;2:349-60.

ISSUE #2: WHAT TO DO?

- Resist temptation as much as possible
- Use some guiding rules
- Remember issue #1 and its “what to do?”

1. Magnitude of difference
2. Statistical significance
3. Pre-existing hypothesis
4. Number of hypotheses
5. Internal consistency
6. External consistency
7. Biologic plausibility

Oxman and Guyatt Ann Intern Med 1992; 116:78-84.

ISSUES #1 AND #2: WHAT TO DO?

Randomized, Double-Blind, Phase II Study of Ruxolitinib or Placebo in Combination With Capecitabine in Patients With Metastatic Pancreatic Cancer for Whom Therapy With Gemcitabine Has Failed

Results

In the intent-to-treat population (ruxolitinib, n = 64; placebo, n = 63), the hazard ratio was 0.79 (95% CI, 0.53 to 1.18; *P* = .25) for OS and was 0.75 (95% CI, 0.52 to 1.10; *P* = .14) for progression-free survival. In a prespecified subgroup analysis of patients with inflammation, defined by serum C-reactive protein levels greater than the study population median (ie, 13 mg/L), OS was significantly greater with ruxolitinib than with placebo (hazard ratio, 0.47; 95% CI, 0.26 to 0.85; *P* = .011). Prolonged survival in this subgroup was supported by post hoc analyses of OS that

Hurwitz et al, JCO 2015;33:4039-47.

ISSUES #1 AND #2: WHAT TO DO?

IDDI
PASSION. SCIENCE. EXPERIENCE.

Shrinkage for clinical trials

Enter the 95% confidence interval for the treatment effect.

Effect

○ log odds ratio ○ log risk ratio ○ log hazard ratio
○ odds ratio ○ risk ratio ◉ hazard ratio
○ risk difference ○ difference of means

Lower

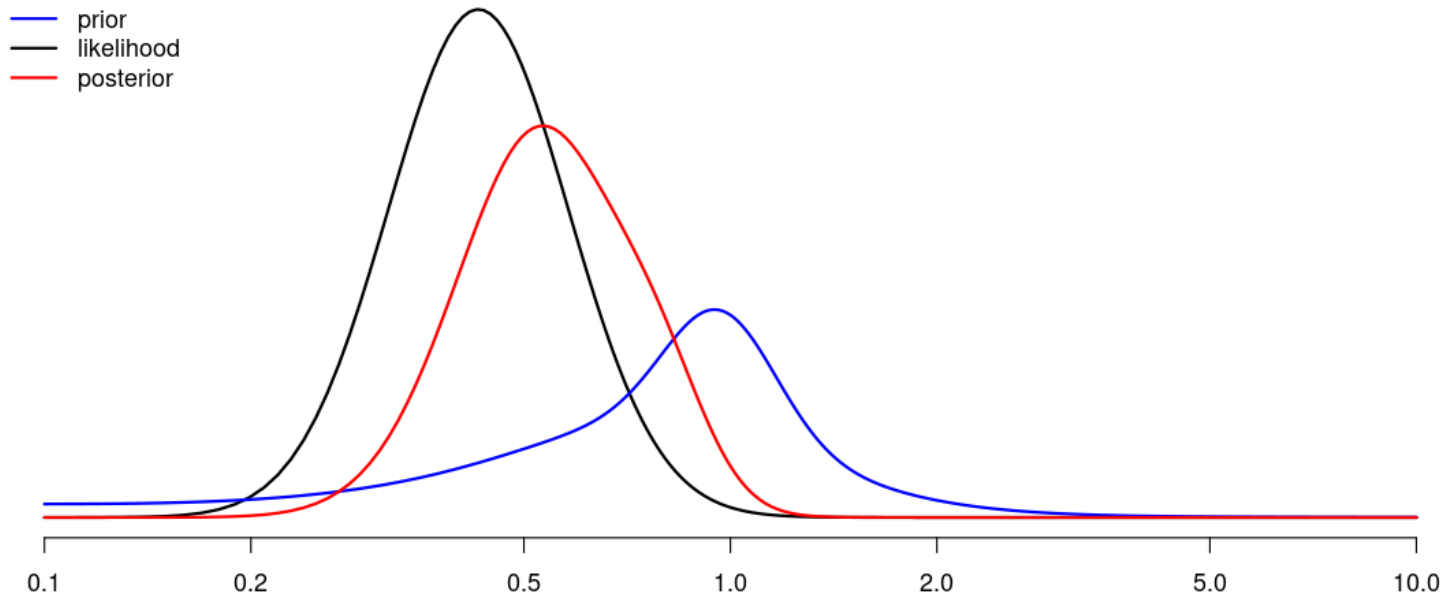
0.26

Upper

0.85

Run

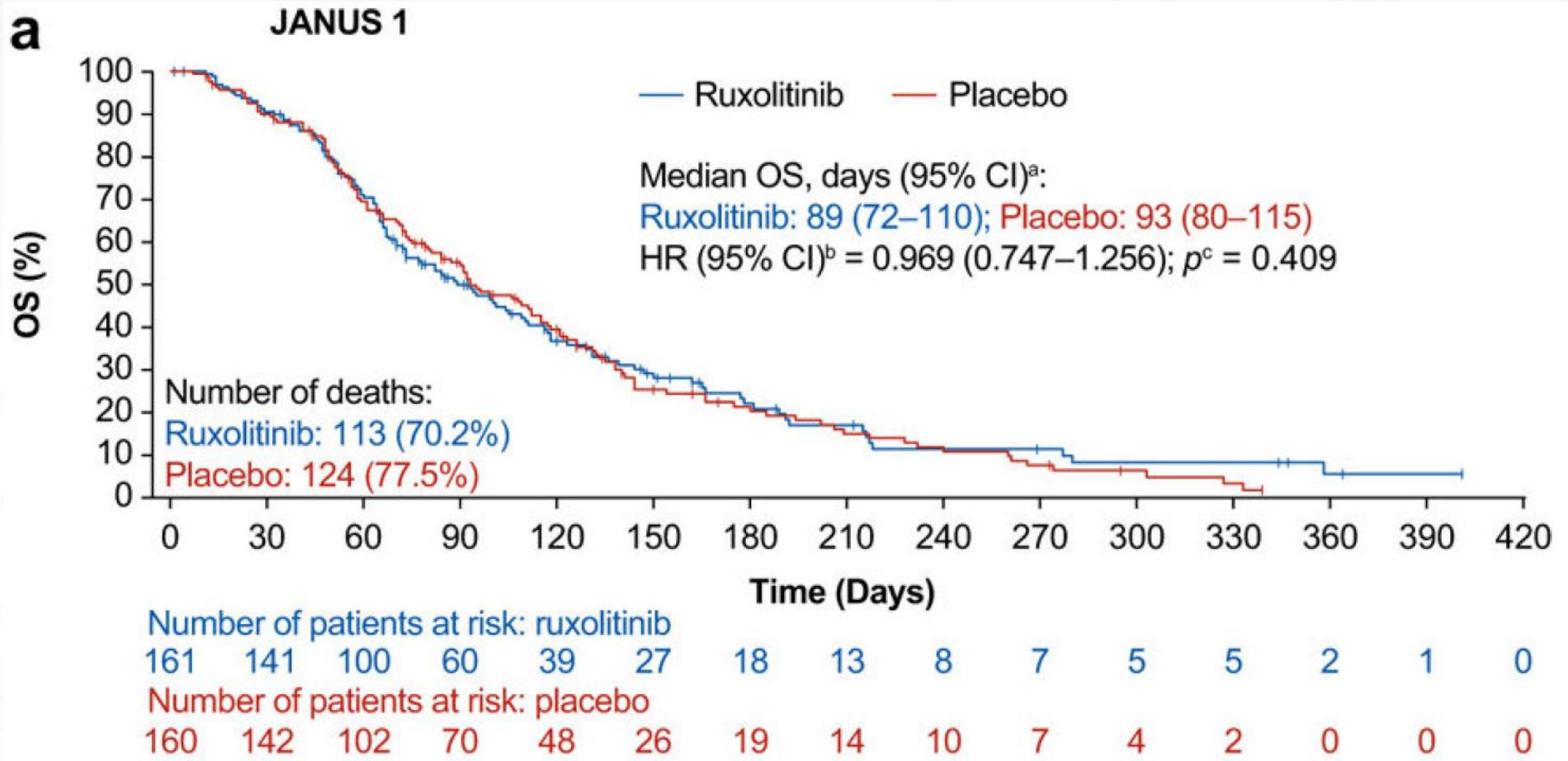
The purpose of this app is to demonstrate the method proposed in the paper "Addressing over-optimism from single RCTs" by van Zwet, Schwab and Greenland in the December 2021 issue of Significance. No rights can be derived from the information offered.



The posterior mean is 0.62 and the 95% credible interval is from 0.33 to 1.01.

https://vanzwet.shinyapps.io/shrinkrct/

ISSUES #1 AND #2: WHAT TO DO?



Hurwitz et al, Invest New Drugs 2018;36:683-95.

ISSUE #3: MISINTERPRETATION OF *P*-VALUES

- **The *P*-value is** the probability of observing a treatment effect at least as extreme as the one observed if there were no true treatment effect (i.e., if the null hypothesis is true) Pr (Data | H_0) 
- **The *P*-value is not**
  - A measure of the magnitude of the treatment effect
  - The probability that the null hypothesis is true Pr (H_0 | Data) 
- Thus, the P-value is not sufficient to quantify the probability that the next trial will be positive

Wasserstein and Lazar, Am Stat 2016;70:129–133.

ISSUE #3: WHAT TO DO?

- Educate yourself on (frequentist) statistics
- Consider a Bayesian framework for early decisions

		Unknown characteristic/truth *Hypothesis*		
		Present (B^c) *Alternative (H_a)*	**Absent (B)** *Null (H_0)*	
Diagnostic *Statistical* **Test Result**	**Positive (A)** *Significant*	**Sensitivity** *Power – true positive*	*Type 1 error (α)*	**PPV**
	Negative *Not significant*	*Type 2 error (β)*	**Specificity** *True negative*	**NPV**

NHST, null hypothesis significance testing; NPV, negative predictive value; PPV, positive predictive value.

- $Pr(Data \mid H_0)$ → $Pr(H_A \mid Data)$ and $Pr(H_0 \mid Data)$

  P-value — Posterior probability of H_A — Posterior probability of H_0

- $Pr(H_A \mid Data) / Pr(H_0 \mid Data)$ is the likelihood ratio or Bayes factor

Ruberg, Clin Pharm Ther 2001;109:1489-98.

P-VALUES VS BAYES FACTOR

P-value	p	0.1	0.05	0.01	0.005	0.001	0.0001	0.00001
Maximum Bayes factor	BFB	1.60	2.44	8.13	13.9	52.9	400	3226
Upper bound of Pr (H_A | Data)	$\Pr^{U}(H_1 \mid p)$	0.62	0.71	0.89	0.933	0.981	0.998	0.9997

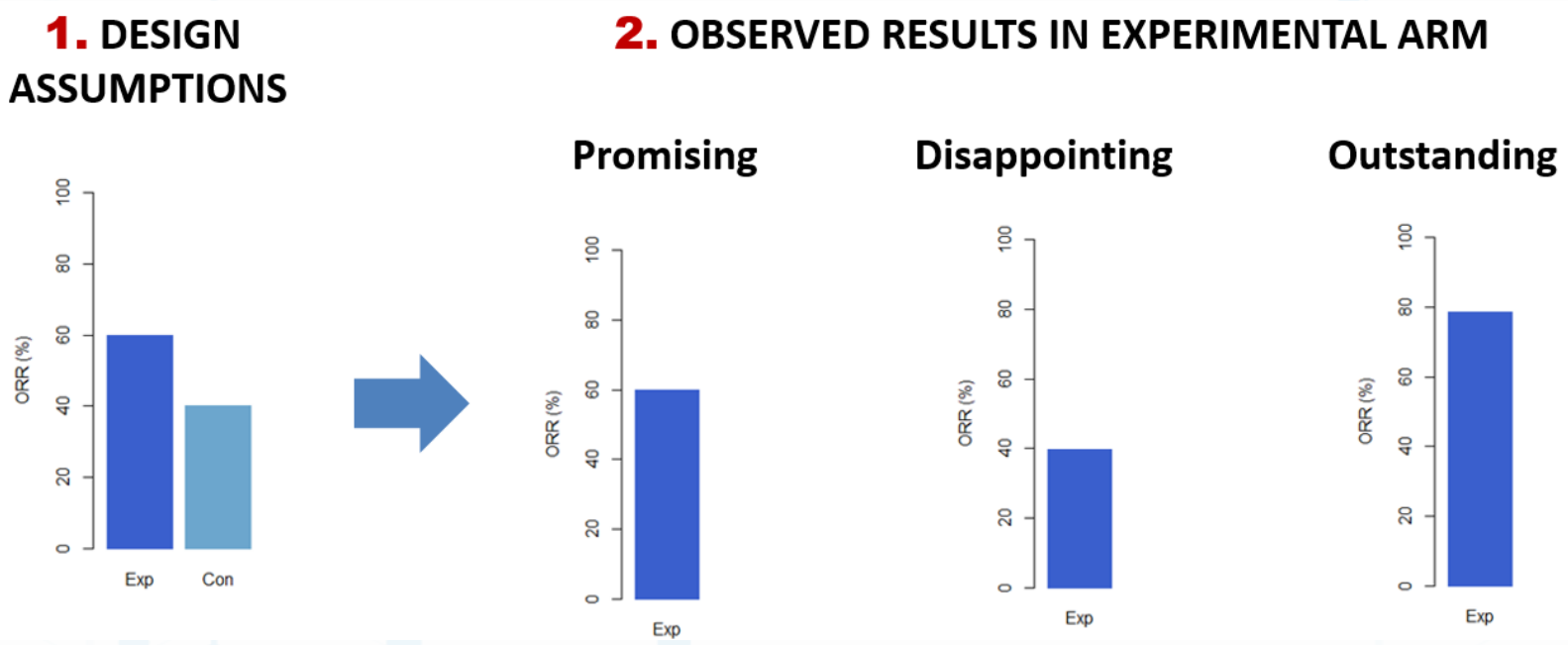
The odds in favor of H_A relative to H_0 are ≈ 2.5:1 for a *P*-value of 0.05 and not the seemingly intuitively 19:1

A P-value <0.005 yields the strength of evidence we often believe we are seeking

Benjamin and Berger, Am Stat 2019;73(Suppl 1):186-91.

ISSUE #4: RELIANCE ON HISTORICAL DATA

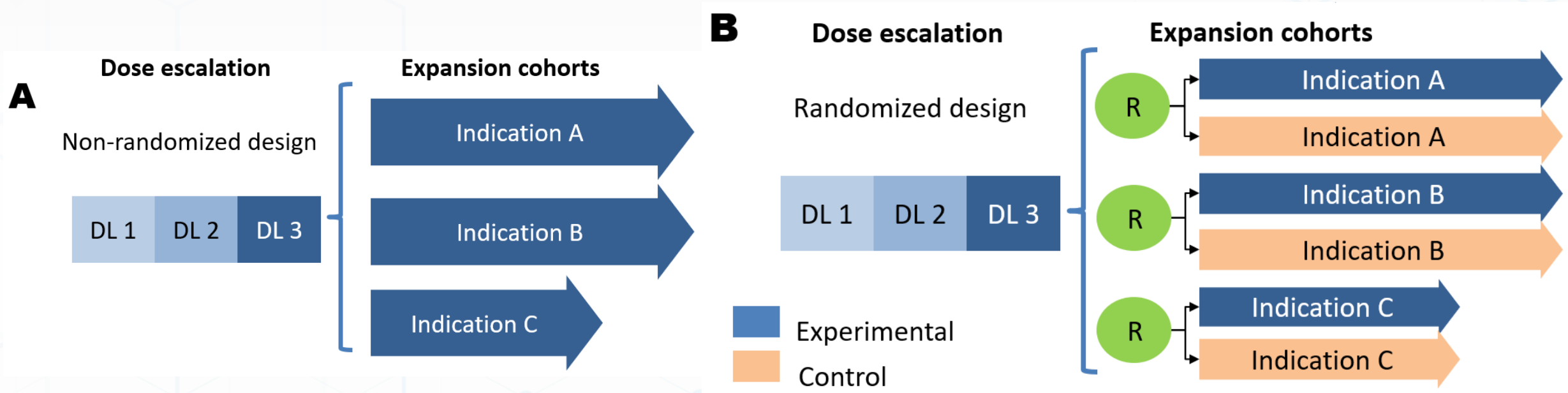
- Several issues combine to make historical data unreliable
  - Selection bias
  - Stage migration
  - Biomarker-defined subsets or uncertainty from other sources



Coart and Saad, Exp Rev Precision Med Drug Dev 2021;6:271-80.

ISSUE #4: WHAT TO DO?

Randomize!



Coart and Saad, Exp Rev Precision Med Drug Dev 2021;6:271-80.

RANDOMIZATION, EVEN 'NON-COMPARATIVE'

2. OBSERVED RESULTS IN EXPERIMENTAL ARM

3. OBSERVED RESULTS IN CONTROL ARM

4. INTERPRETATION

Coart and Saad, Exp Rev Precision Med Drug Dev 2021;6:271-80.

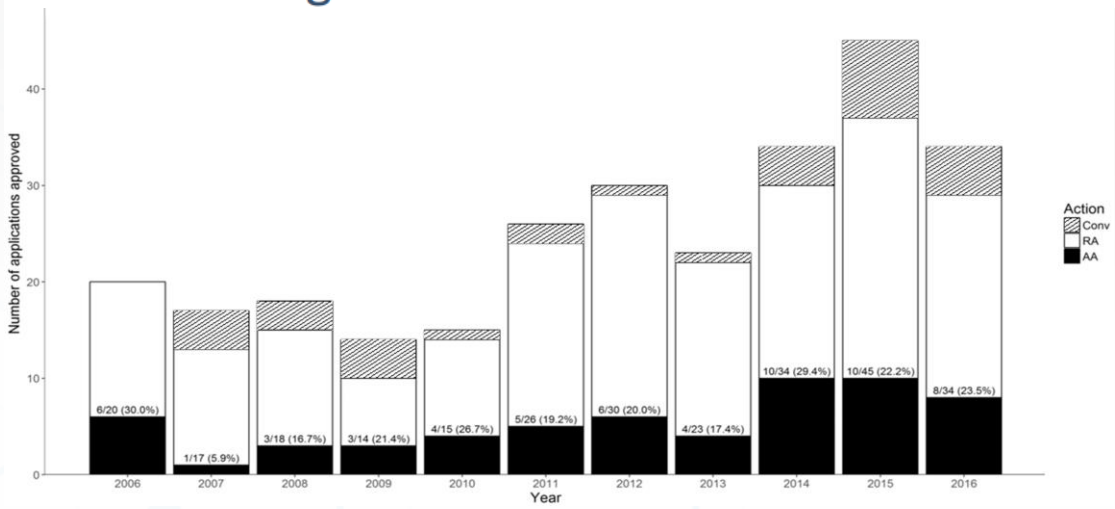
ISSUE #5: TARGETING ACCELERATED APPROVAL

- One of the four expedited programs by the FDA
  - Corresponds roughly to Conditional Marketing Authorization by EMA
  - *Approval based on an effect on a **surrogate endpoint or an intermediate clinical endpoint** that is reasonably likely to predict a drug's clinical benefit*
  - *The sponsor should ordinarily **discuss** the possibility of accelerated approval with the review division during development, supporting, for example, the use of the planned endpoint as a basis for approval and discussing the confirmatory trials, which should usually be already **underway** at the time of approval*
- Oncology drugs account for 66% of all FDA accelerated approvals
  - About 85% in the past decade

https://www.fda.gov/media/86377/download; Fashoyin-Aje et al, N Engl J Med 2022;387:1439-42; GlobalData, Pharma Intelligence Centre (cited in https://www.clinicaltrialsarena.com/comment/oncology-fda-approvals/)

AA: RELATIVE FREQUENCY AND ENDPOINTS

Overview of Oncology and Hematology Drug Approvals at US Food and Drug Administration Between 2008 and 2016

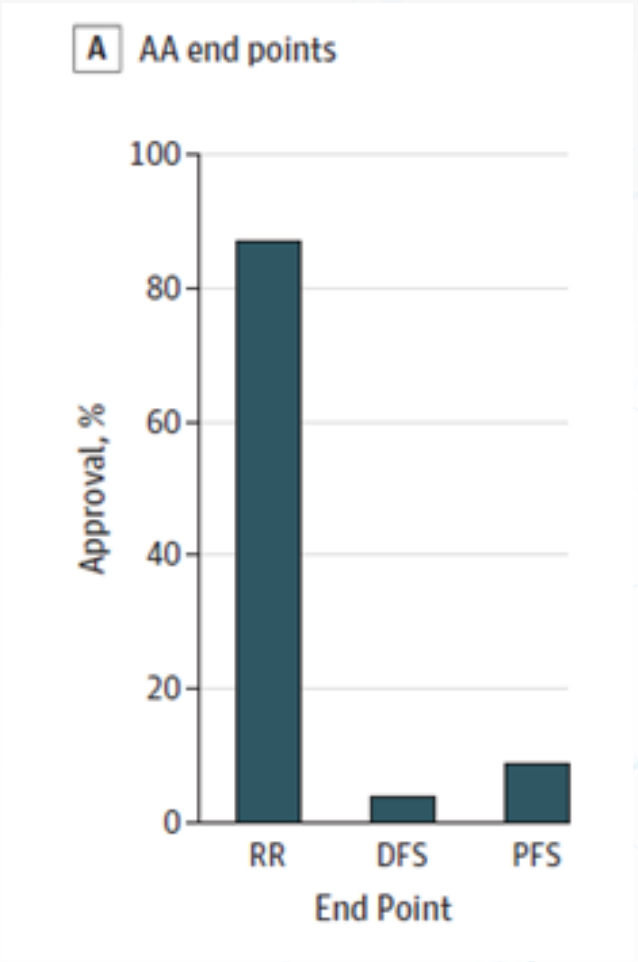


JAMA Oncology | Review

A 25-Year Experience of US Food and Drug Administration Accelerated Approval of Malignant Hematology and Oncology Drugs and Biologics

A Review

Randomized comparative trials supported 26 (28%) of these indications, and single-arm trials accounted for 67 (72%).



Zhou J, et al. J Natl Cancer Inst 2019;111:449-458; Beaver et al, JAMA Oncol 2018;4:849-56.

ISSUE #5: WHAT TO DO?

III. RECOMMENDATIONS

Given the limitations of single-arm trials, a randomized controlled trial is the preferred approach to support an application for accelerated approval. Sponsors can, as appropriate, elect to conduct a single randomized controlled trial to support an accelerated approval and to verify clinical benefit (i.e., follow a "one-trial" approach) or, they can conduct separate trials – one to support the accelerated approval and another, a confirmatory trial, to verify clinical benefit.

Although a randomized controlled trial is the preferred approach, there can be circumstances wherein a single-arm trial is appropriate in the development of a drug for accelerated approval, for example when there are significant concerns about the feasibility of a randomized controlled trial. Careful consideration should be taken in determining whether a single-arm trial is appropriate in a particular clinical and regulatory context. Regardless of the approach under consideration, FDA recommends early discussion with the Agency before initiating and, as appropriate, during the conduct of, a trial(s).

FDA strongly recommends that this trial be well underway, if not fully enrolled, by the time of the accelerated approval action.

To facilitate completion of the confirmatory trial, it may be acceptable to evaluate the drug in the same cancer type but in another line of therapy.

Whether a single trial satisfies the substantial evidence requirement in section 505(d) of the Federal Food, Drug, and Cosmetic Act, should be discussed with FDA early in clinical development, no later than prior to initiating such a trial.

https://www.fda.gov/media/166431/download.

CONCLUSIONS

- Issue #1: “Discount” your early positive results
- Issue #2: Resist temptation, except if results are *really* compelling
- Issue #3: Consider a Bayesian mindset
- Issue #4: Randomize
- Issue #5: Be realistic and consult with agencies early on

THANK YOU !

IDDI
PASSION. SCIENCE. EXPERIENCE.