Two-stage approaches for demonstration of BE.

RBBBD, Amman, 2018.

Anders Fuglsang
Consultant
anfu@fuglsangpharma.com

My background

Former EU assessor, managed the OIP group 2006-2009.

Now consultant working (or worked) for agencies, USP, WHO, companies.

BE trial designs, inspections, audits, fraud detection (incl. algorithms), due diligence.

Two-stage approaches

The idea is simple:

First study a small group of volunteers. Evaluate their data. Use the information collected in the first group to calculate a final sample size. Include the rest of the subjects. Pool the data and conclude.

Ingredients of such a trial:

We need an initial sample size and our desired level of power, ultimately.

One basic issue

Let me share a secret with you:
Regulators don't like alpha's above 5%.
Except when they do.
Jusk ask Helmut Schütz.

After yse for BE.
Per if BE is shown?

This type of design thus allo or sequential testing (i.e. if the first stage fa).

If we apply alpha=5% for both tests, then the overall alpha may be raised above 5%. Therefore we may need to adjust alphas?

Pioneered by Potvin et al. 2008

PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/pst.294

WILEY InterScience® DISCOVER SOMETHING GREAT

Sequential design approaches for bioequivalence studies with crossover designs‡

Diane Potvin[1], Charles E. DiLiberti[2], Walter W. Hauck[3,*,†], Alan F. Parr[4], Donald J. Schuirmann[5] and Robert A. Smith[6]

[1] *Theratechnologies Inc., Montréal, Qué., Canada*
[2] *Barr Laboratories, Inc., Woodcliff Lake, NJ, USA*
[3] *US Pharmacopeia, Rockville, MD, USA*
[4] *GlaxoSmithKline, Inc., Research Triangle Park, NC, USA*
[5] *Center for Drug Evaluation and Research, US Food and Drug Administration, USA*
[6] *Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ, USA*

An absolutely fantastic paper.

Their brilliancy

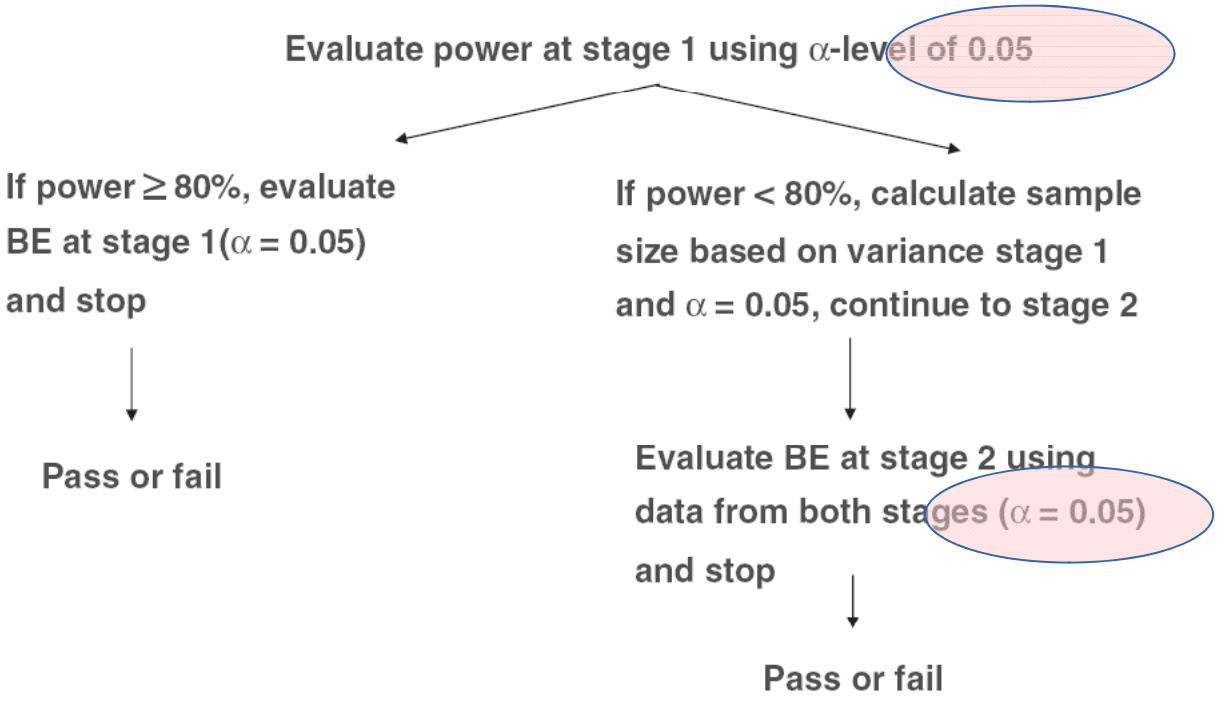
They in invented two/three methods.

They provided a statistical simulation framework for quantifying the type I errors (overall alpha), and the resulting power as function of the initial sample sizes.

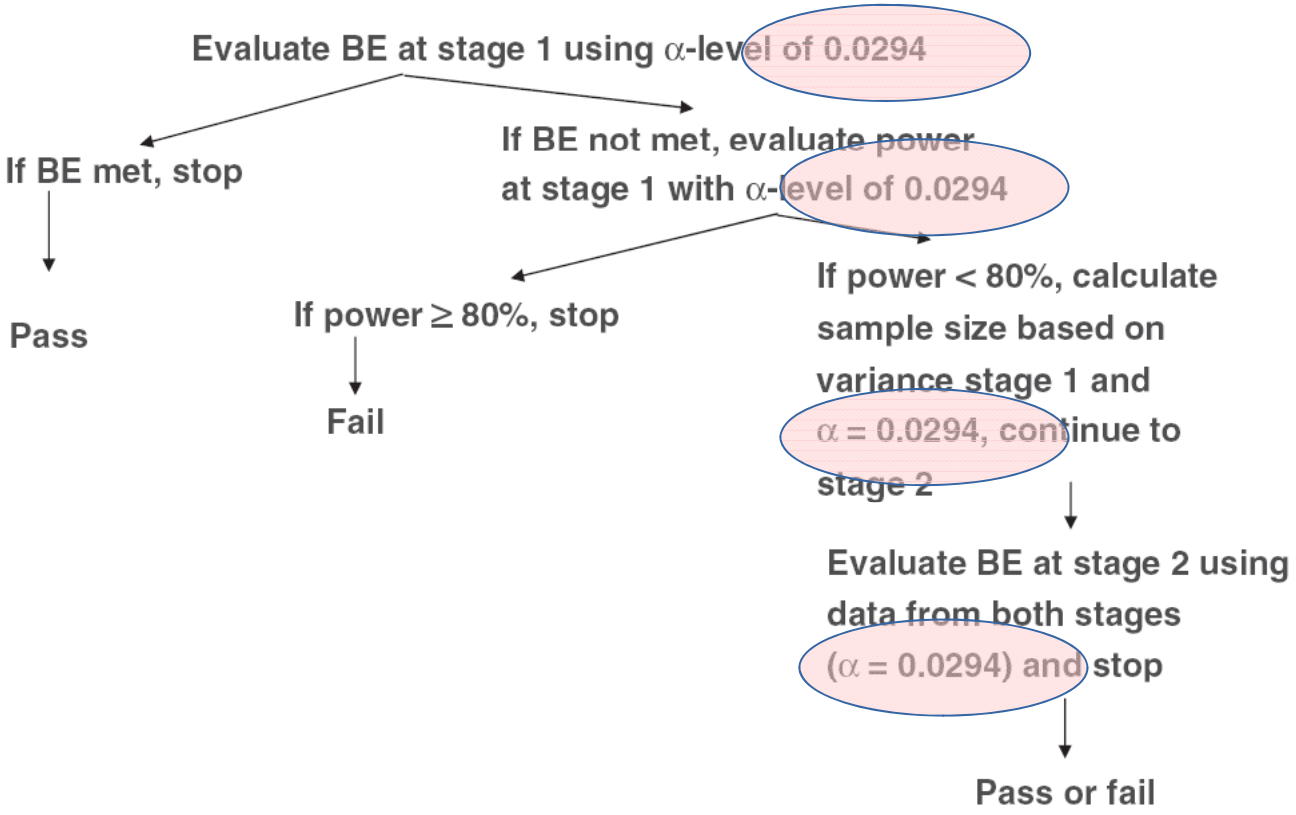
They presented the results for a range of scenarios from N_1=12 to N_1=60, and CV's up to 100%.

Used the 222BE work horse design.

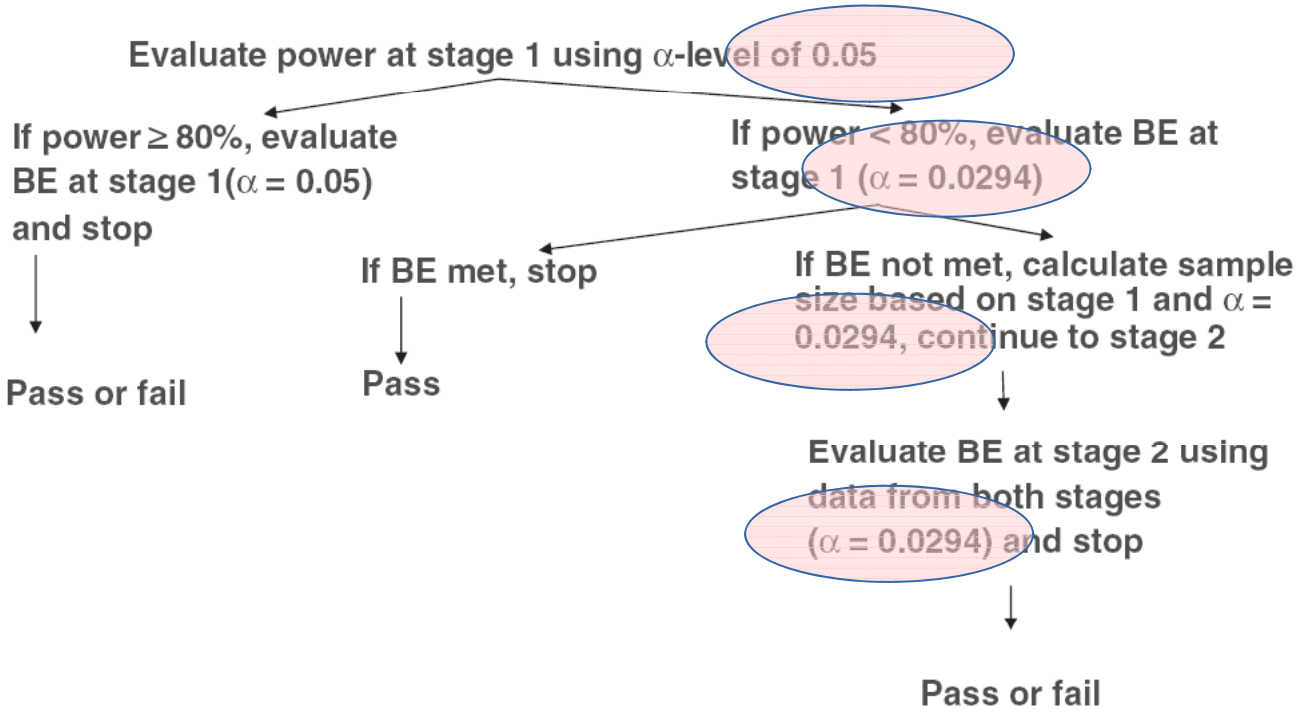
Potvin's method A



Potvin's method B



Potvin's method C



Results

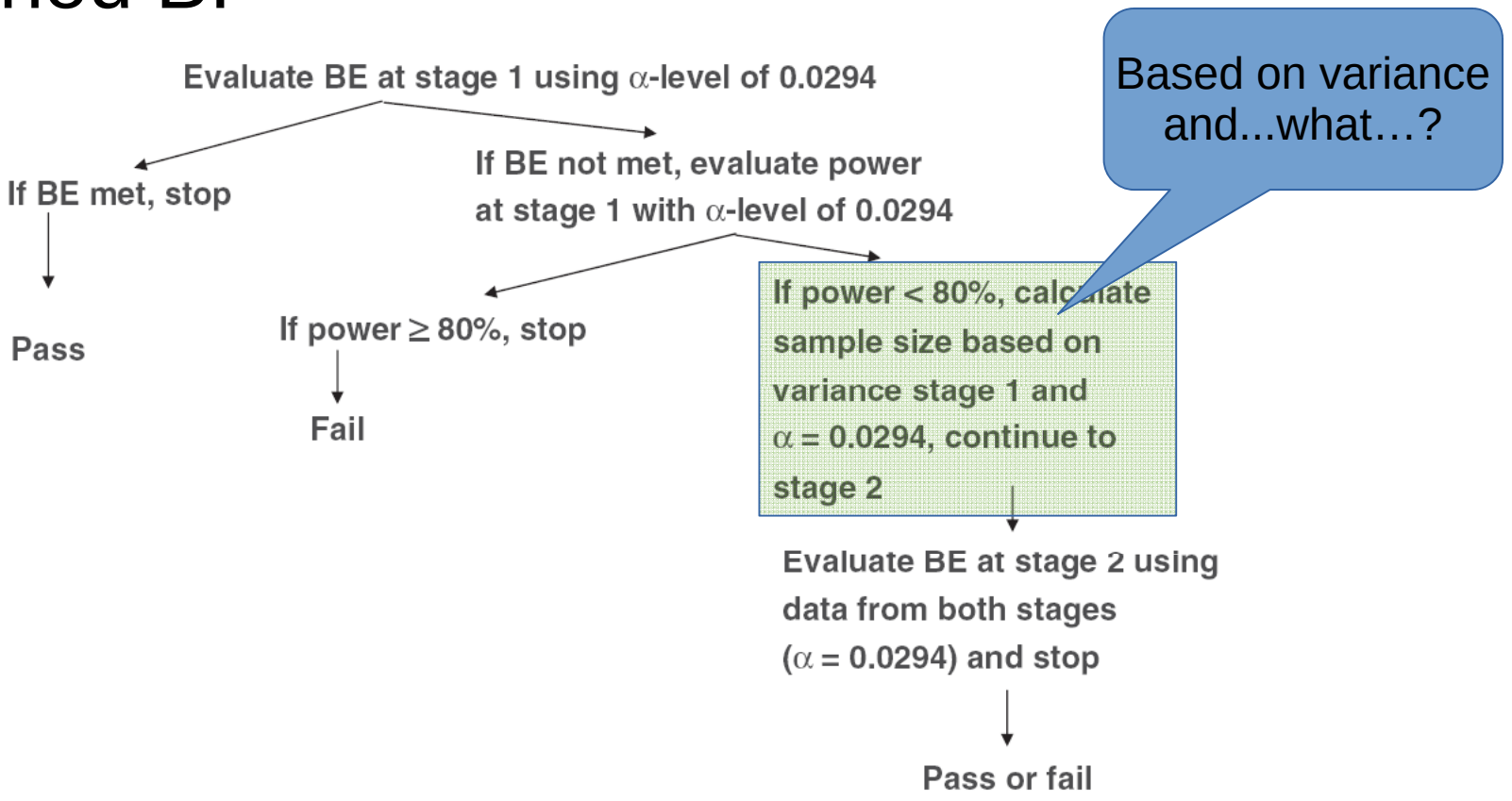
		Estimated type I error rate				Estimated power	
		Method				Method	
Sample size stage 1 (n_1)	Intrasubject CV (%)	A	B	C	D	B	C
12	10		0.0297	0.0496	0.0498	0.9772	0.9890
24	10		0.0294	0.0500	0.0500	0.9999	1.0000
36	10		0.0294	0.0500	**0.0504**	1.0000	1.0000
48	10		0.0292	0.0501	0.0502	1.0000	1.0000
60	10		0.0294	**0.0504**	0.0501	1.0000	1.0000
12	20	***0.0584***	0.0463	**0.0510**	0.0499	0.8429	0.8473
24	20	**0.0505**	0.0320	0.0490	0.0493	0.8810	0.9097
36	20	0.0497	0.0294	0.0499	0.0499	0.9550	0.9750
48	20	0.0500	0.0292	0.0495	0.0497	0.9885	0.9944
60	20	0.0500	0.0297	0.0500	0.0500	0.9973	0.9989
12	30	***0.0575***	0.0437	0.0441	0.0415	0.7857	0.7860
24	30	***0.0550***	0.0475	0.0492	0.0475	0.8305	0.8314
36	30	***0.0523***	0.0397	0.0477	0.0471	0.8379	0.8470
48	30	0.0502	0.0324	0.0494	0.0495	0.8548	0.8873
60	30	0.0498	0.0296	0.0502	0.0499	0.8997	0.9362

Sample size stage 1 (n_1)	Intra-subject CV (%)	Ratio = 1.25				Ratio = 0.95			
		Mean n total (5th, 50th, 95th)		% Studies In stage 2		Mean n total (5th, 50th, 95th)		% Studies In stage 2	
		Method B	Method C	Method B	Method C	Method B	Method C	Method B	Method C
12	10	12.1 (12, 12, 12)	12.0 (12, 12, 12)	3.8	0.9	12.0 (12, 12, 12)	12.0 (12, 12, 12)	0.6	0.2
24	10	24.0 (24, 24, 24)	24.0 (24, 24, 24)	0.0	0.0	24.0 (24, 24, 24)	24.0 (24, 24, 24)	0.0	0.0
36	10	36.0 (36, 36, 36)	36.0 (36, 36, 36)	0.0	0.0	36.0 (36, 36, 36)	36.0 (36, 36, 36)	0.0	0.0
48	10	48.0 (48, 48, 48)	48.0 (48, 48, 48)	0.0	0.0	48.0 (48, 48, 48)	48.0 (48, 48, 48)	0.0	0.0
60	10	60.0 (60, 60, 60)	60.0 (60, 60, 60)	0.0	0.0	60.0 (60, 60, 60)	60.0 (60, 60, 60)	0.0	0.0
12	20	23.2 (12, 22, 40)	23.1 (12, 22, 40)	88.1	80.0	20.6 (12, 18, 40)	20.6 (12, 18, 40)	56.4	53.8
24	20	26.0 (24, 24, 34)	25.4 (24, 24, 34)	34.5	14.2	24.6 (24, 24, 28)	24.4 (24, 24, 24)	8.6	4.4
36	20	36.0 (36, 36, 36)	36.0 (36, 36, 36)	0.7	0.0	36.0 (36, 36, 36)	36.0 (36, 36, 36)	0.1	0.0
48	20	48.0 (48, 48, 48)	48.0 (48, 48, 48)	0.0	0.0	48.0 (48, 48, 48)	48.0 (48, 48, 48)	0.0	0.0
60	20	60.0 (60, 60, 60)	60.0 (60, 60, 60)	0.0	0.0	60.0 (60, 60, 60)	60.0 (60, 60, 60)	0.0	0.0
12	30	47.1 (20, 44, 84)	47.1 (20, 44, 84)	98.0	97.4	46.5 (12, 44, 84)	46.5 (12, 44, 84)	93.1	92.7
24	30	46.9 (24, 46, 72)	46.7 (24, 46, 72)	95.0	90.4	39.9 (24, 38, 70)	39.9 (24, 38, 70)	58.3	56.6
36	30	47.6 (36, 46, 66)	46.5 (36, 46, 66)	81.1	58.0	40.7 (36, 36, 62)	40.5 (36, 36, 62)	29.0	22.7
48	30	51.3 (48, 48, 64)	49.9 (48, 48, 64)	39.2	11.7	48.9 (48, 48, 56)	48.6 (48, 48, 48)	9.4	3.4
60	30	60.3 (60, 60, 62)	60.1 (60, 60, 60)	5.6	0.3	60.1 (60, 60, 60)	60.0 (60, 60, 60)	1.0	0.1
12	100	358.6 (144, 336, 654)	358.9 (144, 336, 654)	100.0	100.0	358.9 (142, 336, 654)	358.9 (144, 336, 654)	100.0	100.0
24	100	358.9 (202, 348, 552)	358.8 (202, 348, 552)	100.0	100.0	358.7 (202, 348, 550)	358.7 (202, 348, 552)	100.0	100.0
36	100	358.8 (230, 352, 512)	358.9 (230, 352, 512)	100.0	100.0	358.9 (230, 352, 512)	358.7 (230, 352, 512)	100.0	100.0
48	100	358.8 (246, 354, 490)	358.7 (246, 354, 490)	100.0	100.0	358.7 (246, 354, 490)	358.9 (246, 354, 490)	100.0	100.0
60	100	358.7 (258, 354, 474)	358.7 (258, 354, 474)	100.0	100.0	358.8 (258, 354, 474)	358.7 (258, 354, 474)	100.0	100.0

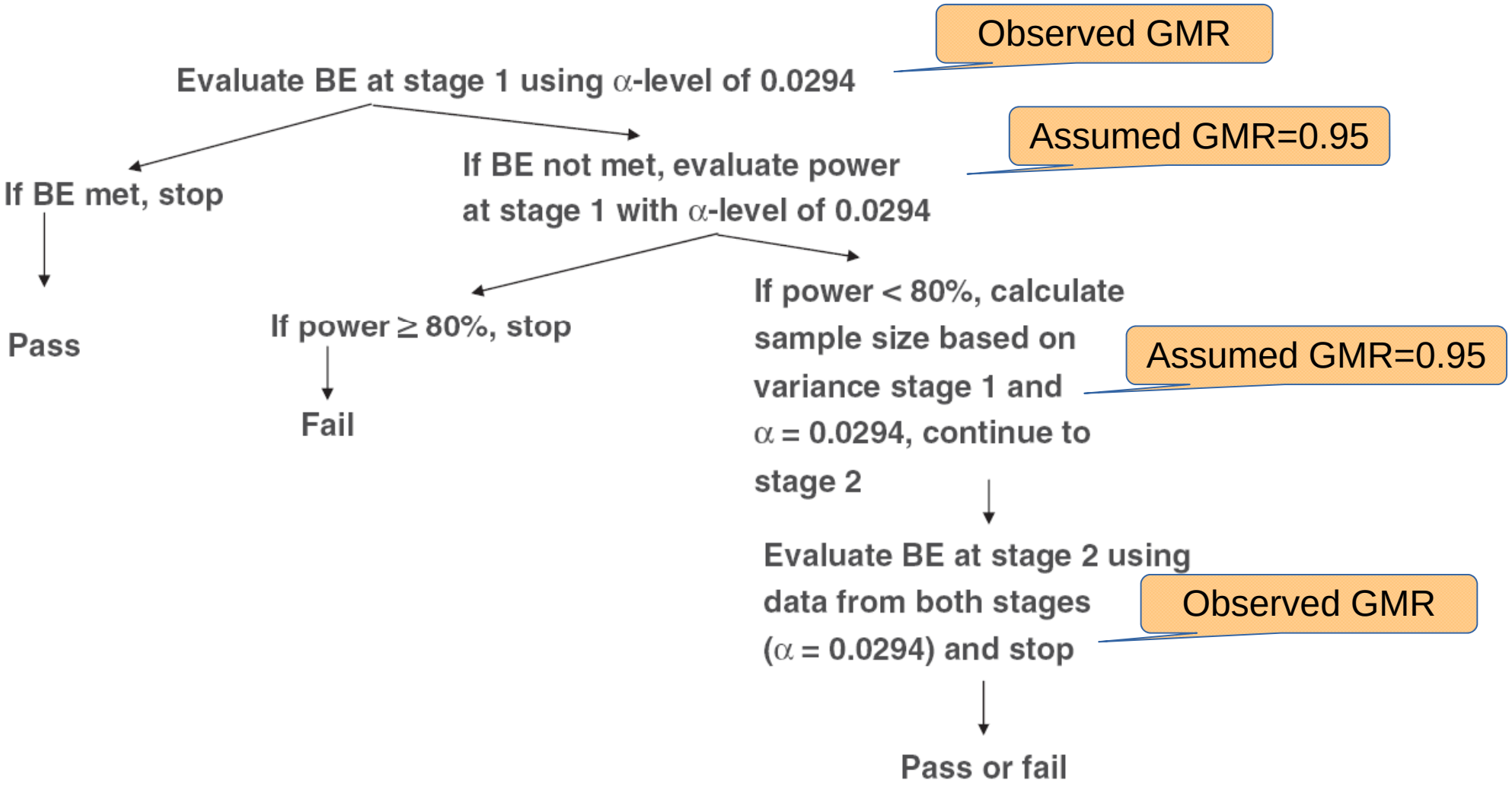
And I thought that was great…

...until I read the paper and the methods again.

E.g. Method B:

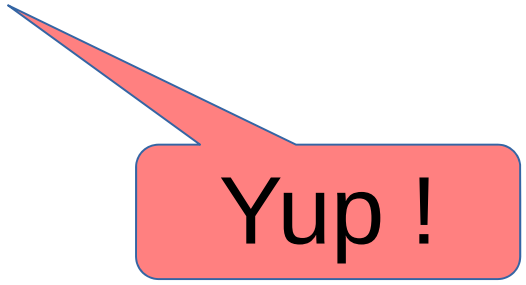


Use observed or assumed GMR ??



Are you saying that….

If we find a stage 1 GMR of 1.89 and a CV of 24%, then we will happily assume GMR=0.95 for the calculation of sample size in stage 2 (if stage 2 is needed) ??



“You cannot be serious!”

(John McEnroe, 1981)

Unfortunately

It turned out that my thinking (and you may well regard this as a pretty general statement) is not as ... thinking.

Numerous people tried in numerous ways to take observed GMR into consideration with no particular luck.

The ... sual assumptions 5... ed GMRs will be farther ... 1 than the true GMR.

a. The sample size grows enormously

b. Often it cannot be calculated at all!

Conclusion

Two-stage trials are pretty handy when there is uncertainty abut the variability and we work with a fixed GMR (means: when we are certain about the GMR (similarity)).

If we are uncertain about the GMR (similarity) then Potvin's two-stage trials have no particular application.

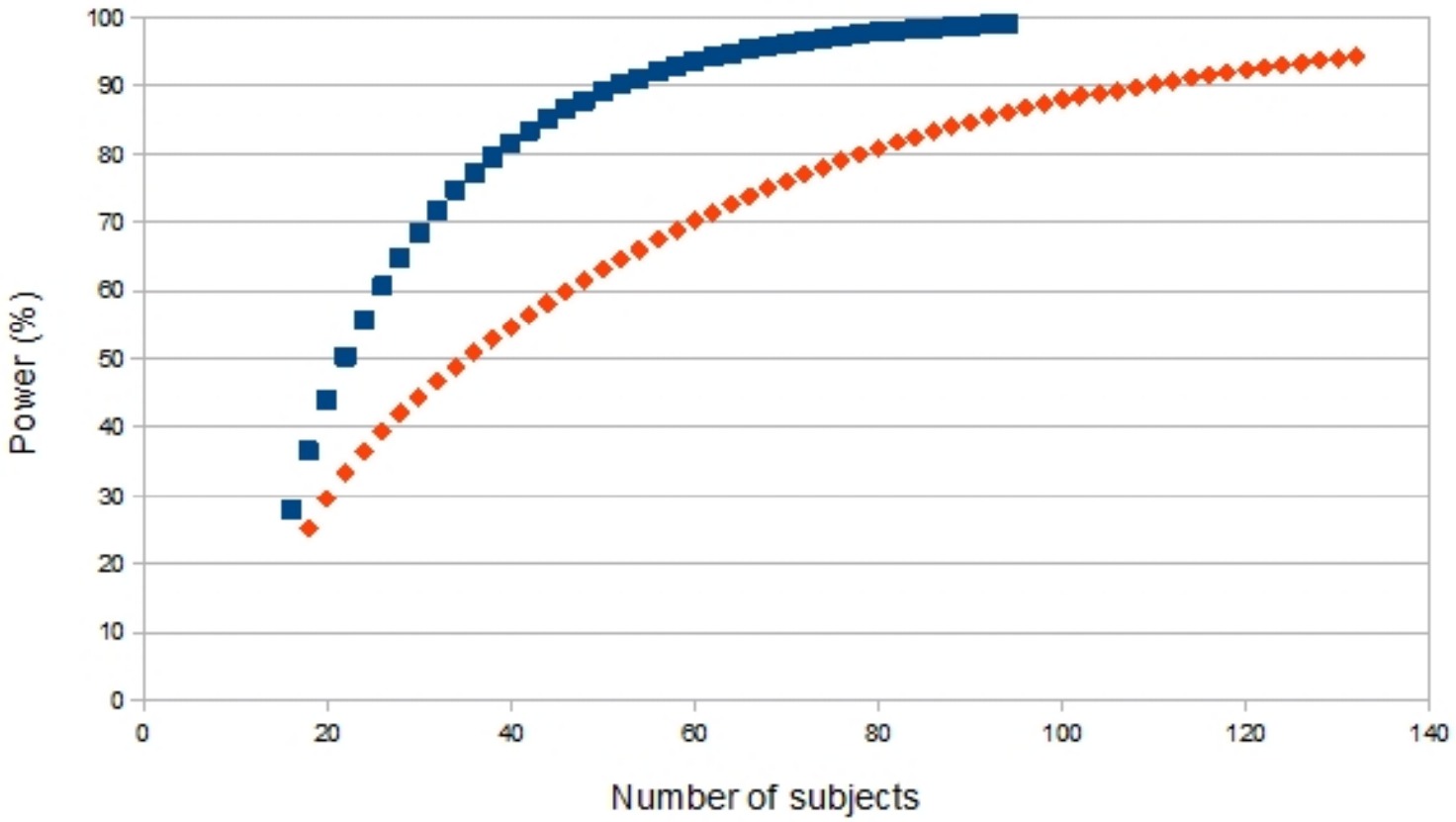
I will word this differently

At a time when there is “too much” uncertainty about the location of the metric of interest (the GMR) we cannot take a qualified decision of basis of its observation (the estimate).

When there is adequate certainty about the location of the GMR we can take a decision about it.

This is very basically why two-stage trials don't work in their present form.

Power example: 2,2,2-design (1 API)



NB!

Blue: T/R=0.95, CV=30%. At N=52 power is 90%.
Red: T/R=0.90, CV=30%. At N=52 power is 65% ! N>100 needed for 90% power.

Numerous potential variations to Potvin's theme

Two-stage trials with two mandatory stages.

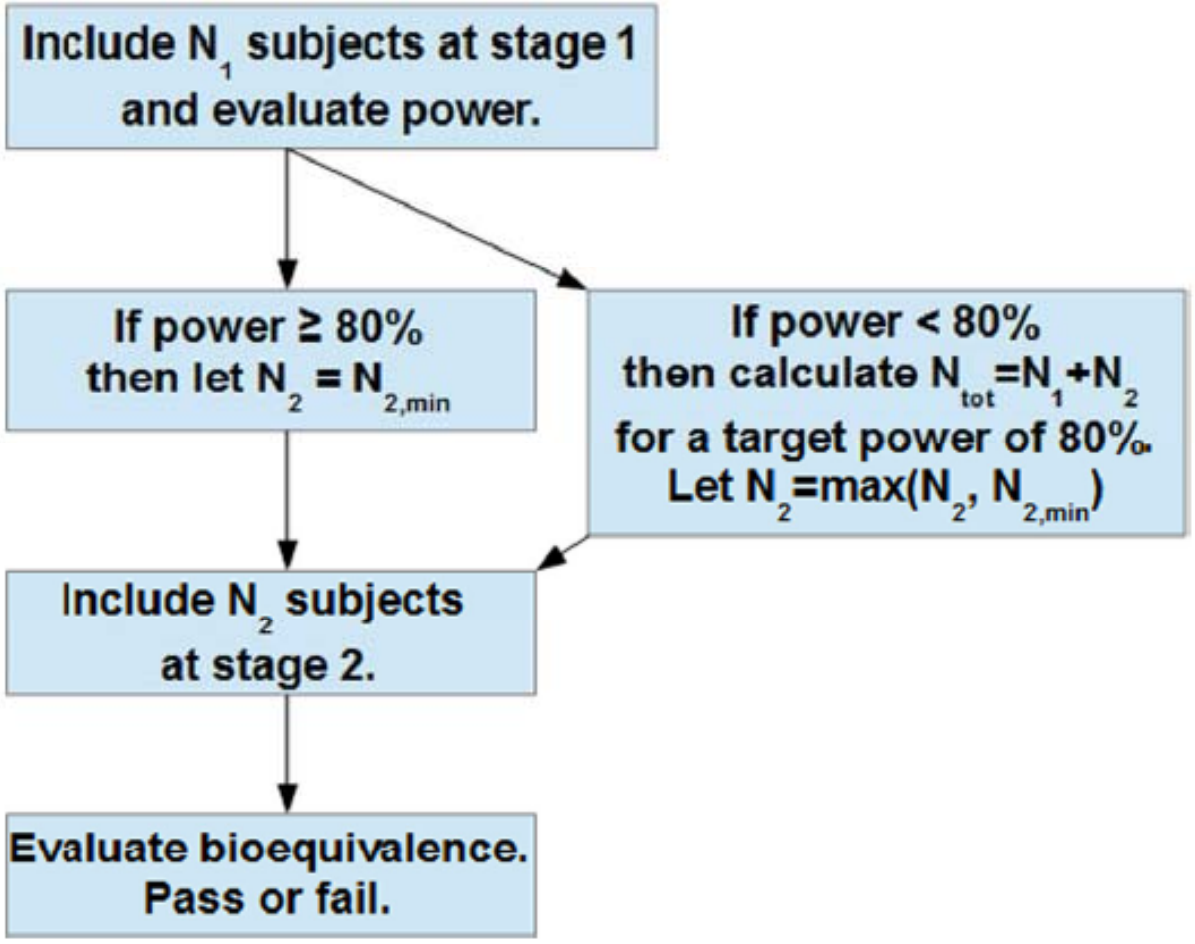
Two-stage trials with futility rules.

Two-stage trials based on parallel designs.

...and more.

Fuglsang 2014

AAPS Journal Journal, Vol. 16, No. 4, July 2014



Fuglsang 2014

AAPS Journal Journal, Vol. 16, No. 4, July 2014

"On comparison of the results obtained here with the results published by Potvin et al., sample sizes, e.g., the average sample sizes, are actually somewha with method E as compared to the variation is re mple size is relative

Blah blah blah.
Leads to a discussion of pilot trials.
Isn't a two-stage trial just a smart way to do a pilot trial with subsequent pivotal trial?

Question

You conduct a 2,2,2-BE pilot study (N=20, say) to learn about comparative product performance before doing a pivotal trial.

Result: You get a **GMR estimate of 0.840** and an apparent **CV=0.287 (28.7%)**.
[Implies: The 90% CI is 72.0%-98.0% for this pilot]

Question: Would you do a pivotal trial with your formulation?

If yes, which CV and point estimate would you use to calculate pivotal sample size?

Question cont'd

Result: *You get a* ***GMR estimate of 0.840*** *and an apparent* ***CV=0.287 (28.7%)****.*
[Implies: The 90% CI is 72.0%-98.0% for this pilot]

If we use GMR=**0.84** then we need **N=414** to get 80% power at CV=0.287 (2,2,2-design).

If we use GMR=**0.90** then we need **N=72** to get 80% power at CV=0.287.

If we use GMR=**0.95** then we need **N=36** to get 80% power at CV=0.287.

What's your proposal?

Question #3

If the first trial is bioinequivalent, should we just stop (and fail)?

Conclusion	Confidence interval (%)
Bioequivalent	
Bioinequivalent (BE not shown)	
Inconclusive (BE not shown)	

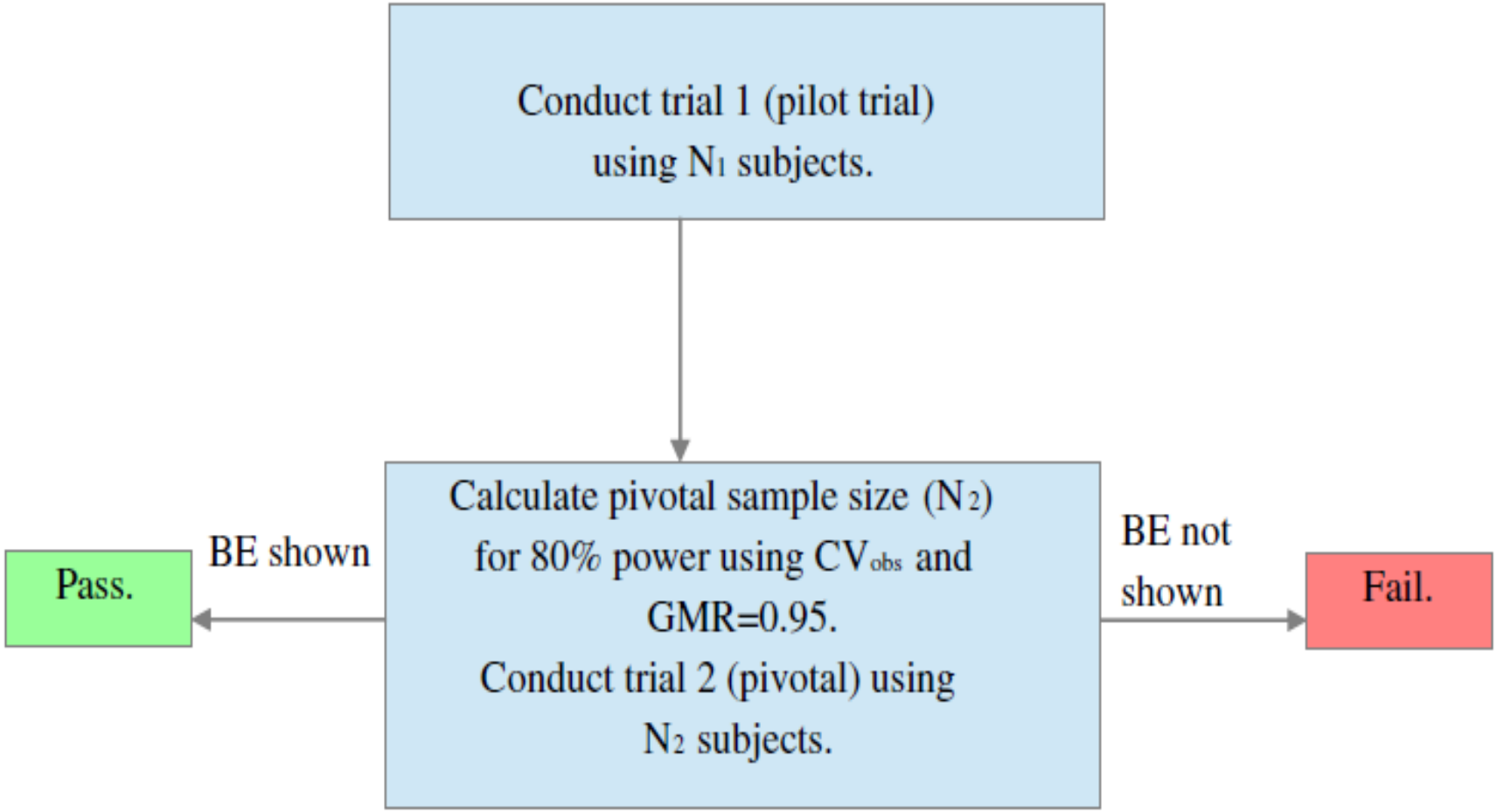
The AAPS Journal (© 2015)
DOI: 10.1208/s12248-015-9744-6

Research Article

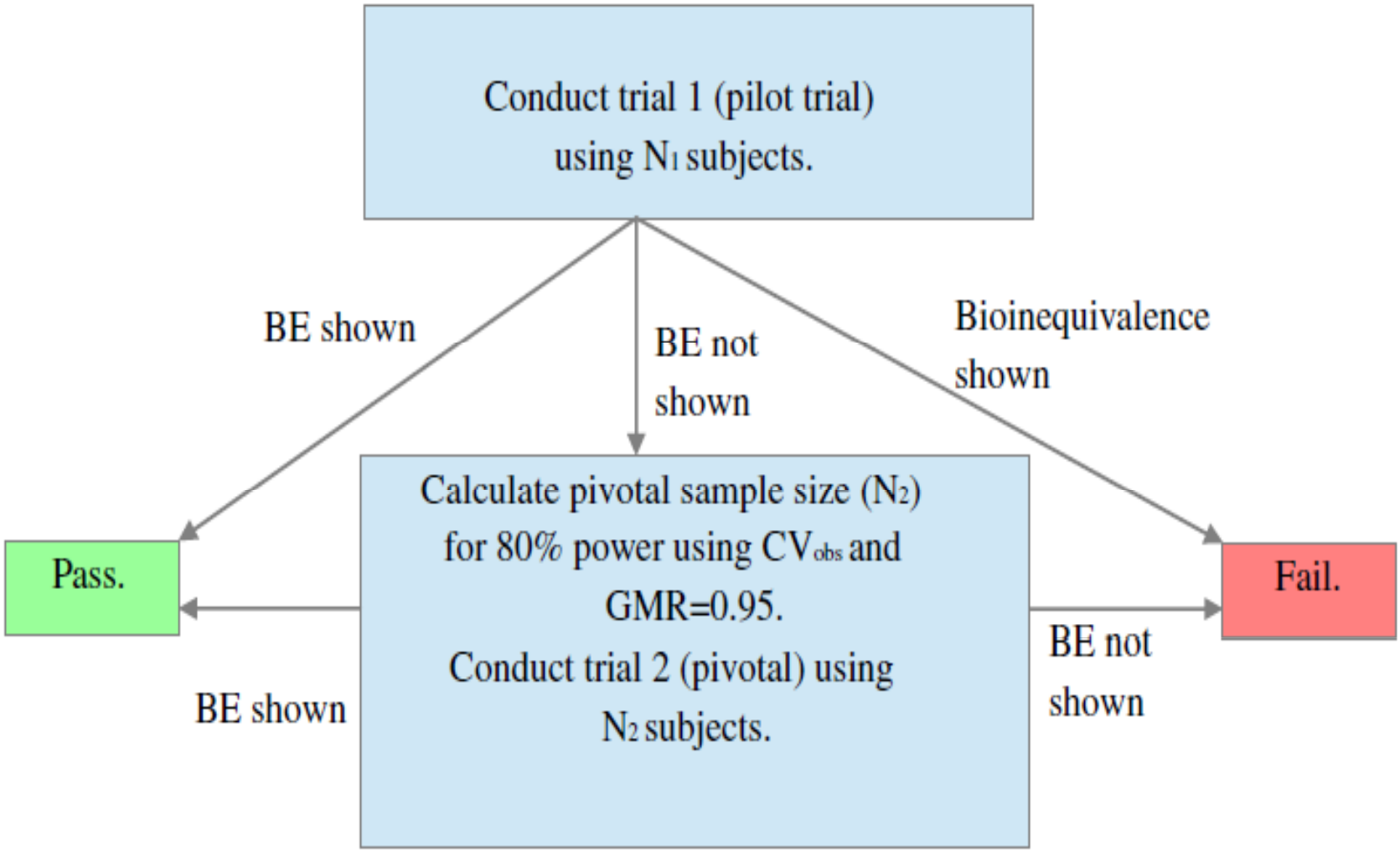
Pilot and Repeat Trials as Development Tools Associated with Demonstration of Bioequivalence

Anders Fuglsang[1,2]

A simple pilot/pivotal trial pair



A more complex pair



The usual methodology

I looked at CV's from 0.1 and upwards.

- 1.000.000 simulations per scenario.
- with and without allowing BE at first trial.
- with and without using the observed PE from the first trial. If the observed PE was not used, the default value of 0.95 was used as in Potvin.
- with and without differentiating between a failed first trial and a bioinequivalent frist trial (failure if inequivalent).

Three important conclusions

- Type I errors easily get inflated if we allow conslusion of BE twice at alpha=5%.
- Sample size often skyrockets when we use the observed GMR from the first trial for planning of the second trial.
- It is not possible to identify a method that consistently gives relatively high power and low type I error rate while keeping the sample size relatively low when GMR is not controlled to 0.95 or better.

It means

- I don't know how to properly and practically apply a pilot trial and use its information if the purpose involves acquiring knowledge of the GMR for any purpose.
  - -When to reformulate vs. when to execute a pivotal trial.
- Pilot trial are reasonable if we know we control the GMR or just wish to know the CV.
- (But Potvin's 2-stage design may be better)

Thanks for listening.
Please get in touch!

anfu@fuglsangpharma.com