PSI webinar – April 23, 2020

Empowering Phase II Clinical Trials to Reduce Phase III Failures

D. De Martini

Università degli Studi di Milano-Bicocca

Ph3 failures: magnitude of the problem

- **Failure rate** in Ph3 CT, is about **45%.**
- 3800 Ph3 trials run worldwide, ≈ 500 patients each.
  - 855,000 patients undergo a failing Ph3 trial.
- $42,000: cost of one patient in Ph3.
  - **$36bn** is spent, yearly, **without achieving the aims** of confirmative trials.

Fair evaluation of Ph3 failure rate

❑ Main reasons of failures are:

- 57-66% : lack of efficacy;
- 9-21% : safety reasons;
- 18-22% : risk-benefit considerations (companies' commercial decision not to file for approval).

➢ **Statistical errors** (type I and type II errors): an ineradicable part of the game:

I. 5-14%: false positive Ph2, translating into Ph3 trials.

II. 10-20%: false negative Ph3 findings ("beta" errors).

These errors give **expected failures** due to lack of efficacy.

❑ **Our target**: rate of **failures due to a lack of efficacy which are not expected** (LNE), i.e. errors in planning.

LNE: numbers

- Estimated rate of **LNE**: ≈ **14%**
- (see "Phase III Failures for a Lack of Efficacy Can Be, in Significant Part, Recovered" paper available at: ssrn.com/abstract=3488251)
- 90% credibility interval: [9%,18%].

- ≈ 1/3 (i.e. 14%/45%) of failures due to **errors in planning**.

❑ This failures translate (yearly) into:

- 270.000 patients that uselessly undergo a Ph3;
- $11bn of waste (costs);
- missed revenues: unknown (often ≈ 10^2 x costs)

- LNE **failures can be recovered** through adequate planning, based on reliable information: -> **enlarging Ph2**.

14% LNE: reasons and countermeasures

- High rate of **underpowered Ph3** trials, due to errors in planning.

❑ Reasons:

I. too optimistic assumptions on the effect size;

II. poor estimation of the effect size.

❑ Countermeasures:

I. avoid assumptions, **use estimates**;

II. **avoid poor estimates**, improve Ph2 information.

- **Enlarging Ph2** can increase the probability of success (if the treatment is effective).

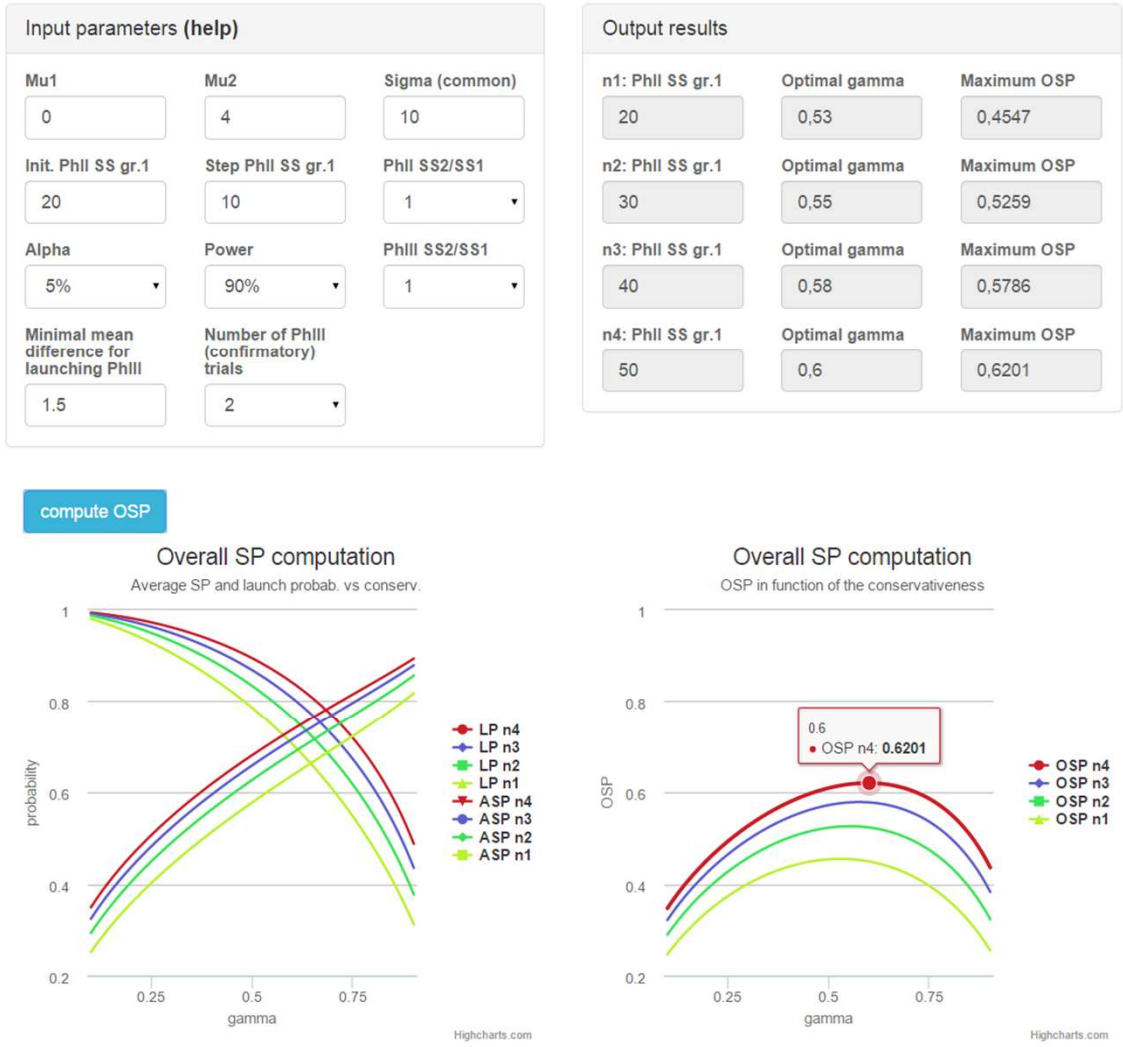
"*We highlight the substantial risk of planning the sample size for confirmatory trials when information is very uninformative*".

(Wang, Hung, O'Neill. Paradigms for adaptive statistical information designs: practical experiences and strategies. Stat.Med. 2012; 31: 3011-3023.)

How large is the benefit by enlarging Ph2?

- One might look at the *overall probability of success* of Ph2 **and** Ph3.
  - See: "Adapting by calibration the sample size of a Ph3 trial on the basis of Ph2 data". Pharm.Stat. 2011; 10(2):89-95.
  - See also: "Success Probability Estimation, with Applications to Clinical Trials". Wiley, 2013.
  - And use sp4ct, at www.sp4ct.com (free)

Dimensioning Ph2: Overall Success Probability of Ph2 *and* Ph3, given the Ph2 sample size



This is a “what happens if” information

Further strategies to reduce failures

- Empowering through enrichment
  (i.e. predictive e. and prognostic e.)
  - **adoption of biomarkers**.

- Allow for greater flexibility in Ph3:
  - **use adaptive designs**;
  - **adoption of surrogate endpoints**.

Biomarkers adoption (BMs)

- Ph3 success rate adopting BMs: ≈ 76% (vs 55%=1-45%).
- Ph2 success rate adopting BMs: ≈ 46% (vs tot 28%).

❑ **Barriers** of BMs:

I. adoptions in few therapeutic areas (e.g. oncology, CV, psychiatry);
II. lack of generalizability;
III. need of identifying and validating new BMs.

❑ **Facts**:

I. BMs are adopted in ≈ 5% of trials (76% and 46% come from 5% of trials);
II. this 5% rate seems hard to grow quickly.

Adaptive Designs (ADs)

- Ph3 success rate with ADs: unknown.
- A wider application in Ph2 of early stopping ADs would allow saving resources for empowering other Ph2 trials.
- A wider application in Ph3 of early stopping and of ss re-estimation ADs would reduce waste.

❑ **Barriers** of Ads in Ph3:

I. Regulatory Agencies: concerns about type I error increase;

II. oxymoron: adaptively searching, in Ph3, for the right population, the right dose, or the right endpoint in confirmative trials may constitute an oxymoron.

❑ **Facts**:

I. ADs adoption rate in Ph3 is < 1%;

II. adoption rate did not increase in the last 10 years.

Surrogate Endpoints (SEs)

- Ph3 success rate with SEs: unknown.
- When SEs are adopted, trials seem presenting a high success rate (infectious diseases) (Hay et al. Clinical development success rates for investigational drugs. Nature Biotechnology 2014; 32(1): 40-51.)

❑ **Barriers** of SEs in Ph3, skepticism:

I. "it may be that trials that attempt to evaluate the effectiveness of biomarkers are more likely to fail". (Wong et al. Estimation of clinical trial success rates and related parameters. Biostatistics 2018).

II. "biomarkers may fail to provide reliable evidence about the benefit-to risk profile of interventions". (Fleming and Powers. Biomarkers and surrogate endpoints in clinical trials. Statistics in Medicine 2012; 31(25): 2973-2984).

❑ **Facts:**

I. validation of SEs is very rigorous.

Concerns about Ph2 enlargement

- **Barriers**:

I. Not all Ph2 trials give estimates of Ph3 effect size; for example, in cancer trials (40% of the total), Ph2s measure tumor dimensions, where Ph3s focuse on survival (note that these two outcomes are often correlated).

II. Economic evaluations: longer development times, higher costs.

- **Facts**:

I. New treatments, small companies: 74% of Ph2 are run by emerging companies.

II. 60% of these Ph2 are unpartnered: a partnership could support adequate Ph2.

III. Ph2 is often small: on average ¼ of Ph3 (ss ≈ 125 vs 500).

IV. Good Ph3 planning needs: Ph2 ss ≥ 2/3 of the ideal Ph3 ss.

Conclusions

- Expanding Ph2, and conservatively plan Ph3, seems not only ethical but also mandatory.
- Is, expanding Ph2, valuable?
  - With one treatment: enlarging Ph2 reduces Ph3 underpowering, increases the probability of success, and the expected profit.
  - With a set of treatments: when the resources are limited, not all Ph2s can be expanded; on average, one in three Ph2 trials can enlarged and launched.

Perspectives with a set of treatments

- Pharma company: simulate profit behavior, to decide which drug developments merit launching.
- Public Health: more complex: the set of treatments can be quite large; a metric for ranking the potential impact on public health should be established; priority setting studies may be helpful.

❑ In general: when enlarging Ph2s, less Ph2s launched, less Ph3s launched, higher Ph3 success rate, and lower number of successful Ph3s. … however, the treatments developed, that succeed with a higher rate, might be of higher quality.

❖ To conclude, there is not a unique answer for enlarging Ph2, case-by-case solutions should be adopted.

End

- Thanks for attending
- ➢ The full paper, and related References, is available here: https://onlinelibrary.wiley.com/doi/10.1002/pst.1980