

On the Cost of Running Two-by-Two Blind Trials

Sylvain Chassang

Erik Snowberg

Princeton University

Caltech and NBER*

July 25, 2015

Motivation

Two-by-two blind trials (Chassang et al., 2012, 2015), illustrated in Figure 1, randomize both treatment status, and the patients' likelihood of receiving treatment. This allows medical researchers to measure pure treatment effects, Placebo or behavioral effects, and interaction effects between treatment and behavior. When there are interaction effects between treatment and behavior, the benefits of two-by-two trials are clear: they improve our measures of the value added of drugs, provide researchers with a clear signal that behavior needs to be further investigated, leading to more efficient adoption of new treatments.

This note discusses the costs of blind trials when there are no interaction effects between treatment and behavior. We argue quantitatively that for plausible implementations of blind trials these costs are small, or non-existent. There are two reasons for this: first, the data of two-by-two trials can be re-aggregated in a single estimate of the usual treatment effect; second, drop-out rates tend to be low in high likelihood of treatment arms.

*Chassang: chassang@princeton.edu, Snowberg: snowberg@caltech.edu.

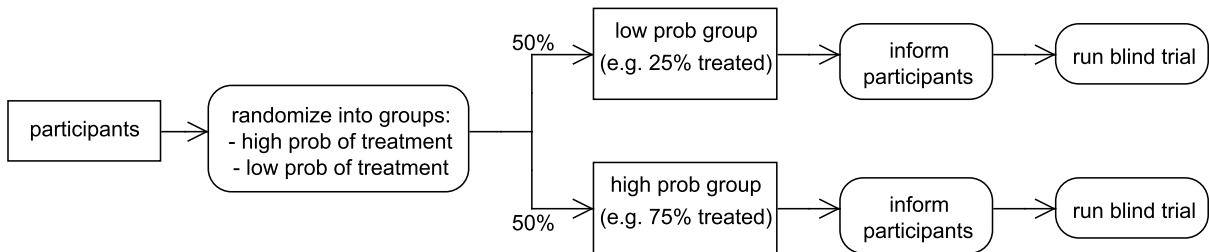


Figure 1: a two-by-two trial

1 Re-aggregating Trial Data

The main potential cost of two-by-two trials is that they use non-symmetric treatment and control groups which in principle comes at some cost of power, equivalent to a small proportional loss in sample size. A countervailing force is that higher than 50-50 likelihood of treatment lowers drop-out rates.

Homoscedastic treatment effects. The forces at work are best explained in the context of the standard homoscedastic treatment effect model. Let $p \in [0, 1]$ denote a patient's probability of being assigned to the treatment group, and τ denote her treatment status. Assuming no interaction effects between treatment and behavior, outcomes Y_i take the form

$$Y_i = \mu\tau_i + \gamma_p + \sigma\varepsilon_i, \quad (1)$$

where γ_p is a pure behavioral or placebo effect, allowed to depend on probability of treatment p , ε is distributed according to a standard normal $\mathcal{N}(0, 1)$, and $\sigma > 0$ is fixed. Extending the analysis to heteroscedastic environments with known variances is immediate.

Precision from a standard blind trial. Consider a standard blind trial with likelihood of treatment p and drop-out rate $1 - \rho_0 \in [0, 1]$. Let N denote the initial sample size, and \bar{N}_τ the set of patients with treatment status τ completing the trial. The Wald estimator of

treatment effect μ is

$$\hat{\mu}_0 \equiv \frac{1}{\rho_0 p N} \sum_{i \in \bar{N}_1} Y_i - \frac{1}{\rho_0 (1-p) N} \sum_{i \in \bar{N}_0} Y_i, \quad (2)$$

$$\sim \mathcal{N} \left(\mu, \frac{\sigma_0^2}{N} \right) \quad (3)$$

with $\sigma_0^2 \equiv \frac{\sigma^2}{\rho_0} \left(\frac{1}{p} + \frac{1}{1-p} \right)$. Keeping drop-out rate ρ_0 fixed, the variance of $\hat{\mu}$ is minimized by setting $p = 1 - p = \frac{1}{2}$. In that case $\sigma_0^2 = \frac{4\sigma^2}{\rho_0}$.

Re-aggregating data from two-by-two trials. Under two-by-two trials, patients are randomly assigned to a high and low probability of treatment group with respective probability of treatment denoted p_H and p_L . For simplicity, we assume throughout that $N/2$ patients are respectively assigned to the high and low probability of treatment groups. We denote by ρ_L and ρ_H patient-retention rates in each group. We denote by \bar{N}_τ^G the set of patients in subgroup $G \in \{L, H\}$ with treatment status $\tau \in \{0, 1\}$ completing the trial.

Because the high a low probability of treatment groups may have different Placebo effects, naïvely pooling together the data from the two sub trials and using it to compute a standard Wald estimator generates a biased estimator of treatment effects. Instead we form an estimator for each sub group:

$$\hat{\mu}_L \equiv \frac{2}{\rho_L p_L N} \sum_{i \in \bar{N}_1^L} Y_i - \frac{2}{\rho_L (1-p_L) N} \sum_{i \in \bar{N}_0^L} Y_i \sim \mathcal{N} \left(\mu, \frac{\sigma_L^2}{N} \right),$$

$$\hat{\mu}_H \equiv \frac{2}{\rho_H p_H N} \sum_{i \in \bar{N}_1^H} Y_i - \frac{2}{\rho_H (1-p_H) N} \sum_{i \in \bar{N}_0^H} Y_i \sim \mathcal{N} \left(\mu, \frac{\sigma_H^2}{N} \right),$$

with

$$\sigma_L^2 \equiv \frac{2\sigma^2}{\rho_L} \left(\frac{1}{p_L} + \frac{1}{1-p_L} \right) \quad \text{and} \quad \sigma_H^2 \equiv \frac{2\sigma^2}{\rho_H} \left(\frac{1}{p_H} + \frac{1}{1-p_H} \right).$$

Estimators $\hat{\mu}^H$ and $\hat{\mu}^L$ are unbiased and for any weight $\lambda \in [0, 1]$ can be aggregated into

unbiased estimator

$$\widehat{\mu}_{TbT} \equiv \lambda \widehat{\mu}_L + (1 - \lambda) \widehat{\mu}_H \sim \mathcal{N} \left(\mu, \frac{\sigma_{TbT}^2}{N} \right).$$

where $\sigma_{TbT}^2 = \lambda^2 \sigma_L^2 + (1 - \lambda)^2 \sigma_H^2$. The variance of $\widehat{\mu}_{TbT}$ is minimized by setting $\lambda = \frac{\sigma_H^2}{\sigma_H^2 + \sigma_L^2}$, which yields

$$\sigma_{TbT}^2 \equiv \frac{\sigma_L^2 \sigma_H^2}{\sigma_L^2 + \sigma_H^2}.$$

If $p_H = p_L = .5$, variance coefficient σ_{TbT}^2 coincides with σ_0^2 . Note that optimal weights depend only of attrition rates ρ_L and ρ_H , as well as treatment probabilities p_L and p_H .

2 Quantitative Evaluation

In the absence of interaction effects, the cost of running a two-by-two trial is well measured by the power loss

$$\mathcal{L} \equiv \frac{\sigma_{TbT}^2}{\sigma_0^2} - 1.$$

Loss \mathcal{L} corresponds to the percentage increase needed in sample size N so that $\widehat{\mu}_0$ and $\widehat{\mu}_{TbT}$ have the same standard deviation.

Two-forces affect loss \mathcal{L} : first, some statistical power is lost from the fact that treatment and control groups in each subgroup of the two-by-two trial are asymmetric; second, some statistical power may be gained (resp. lost) if one of the subgroups of the two-by-two trial has higher (resp. lower) retention rate than ρ_0 .

We compute plausible losses as follows. We consider two-by-two trials such that $p_L = 1/2$ (i.e. the low probability of treatment subgroup coincides with a standard blind trial), and consider $p_H \in \{\frac{2}{3}, \frac{3}{4}\}$. Retention rates are gauged from the data of Chassang et al. (2015): blind trials with a probability of treatment $p = \frac{1}{2}$ have a median retention rate equal to 75%; blind trials with a probability of treatment $p \geq \frac{2}{3}$ have a median retention rate equal to 85%.

This leads us to consider losses for values $\rho_0 = \rho_L = .75$, and $\rho_H \in \{.75, .80, .85\}$. Losses \mathcal{L} are summarized in Table 1.

$\rho_H \backslash p_H$	$\frac{2}{3}$	$.7$	$\frac{3}{4}$
$.75$.059	.087	.14
$.80$.027	.055	.11
$.85$	-.003	.025	.08

Table 1: power loss \mathcal{L} as a function of parameters p_H and ρ_H .

The main quantitative take-away is that for values of p_H in $\{2/3, .7\}$ the loss of power is small, and in fact negative when the retention rate ρ_H is high enough. This leads us to the conclusion that there is almost no downside to using a two-by-two trial instead of a standard blind trial.

References

- Chassang, Sylvain, Erik Snowberg, Ben Seymour, and Cayley Bowles, “Accounting for Behavior in Treatment Effects: New Applications for Blind Trials,” *PLOS One*, 2015.
- , Gerard Padro I Miquel, and Erik Snowberg, “Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments,” *American Economic Review*, 2012, *102* (4), 1279–1309.