Zusha! Additional Calculations

Dan Brown

January 2018

1 Statistical Power Calculations

In this Section, we compute power calculations for each of the four studies: Habyarimana & Jack 2011, Habyarimana & Jack 2015, and the upcoming RCTs in Tanzania and Uganda.

1.1 Assumptions for power calculations

1.1.1 Cross-Study Assumptions

We make the following cross-study assumptions.

- There are no reliable covariates that predict the incidence of accidents, and so the study uses a simple two sample test.
- That two sample test is to compare proportions (the accident rate).
- There is no cluster structure to the data generating process predicting accidents.
- The study uses a 95% significance level to detect significant effects.

1.1.2 Tanzania RCT

Additional assumptions:

- Control group has 3,197 participants, treatment group has 2,789.
- Initial accident rate of 0.032 (the preintervention accident rate for the control group, as discussed in conversations with the Zusha! team).

1.1.3 Uganda RCT

Additional assumptions:

- Control group has 3,925 participants, treatment group has 3,950.
- Initial accident rate of 0.119 (the preintervention accident rate for the control group, as discussed in conversations with the Zusha! team).

1.1.4 Habyarimana & Jack 2011

Additional assumptions:

- Control group has 1,006 participants, treatment group has 1,155.
- Initial accident rate of 0.061 (the annualised insurance claim rate for the control group, see Table 1 in Habyarimana & Jack 2011).

1.1.5 Habyarimana & Jack 2015

Additional assumptions:

- Control group has 3,852 participants, treatment group has 7,885.
- Initial accident rate of 0.079 (the preintervention accident rate for the control group, see Table 3 in Habyarimana & Jack 2015).

1.2 Results

1.2.1 Power Calculations

- The results are presented in Table 1.
- Power calculations depend on the postulated true treatment effect size. We have calculated the power of the studies for three assumptions about the true treatment effect: that the stickers decrease accident rates by 20%, 10% or 5%.
- The statistical power of the study, for say a 20% postulated true treatment effect, is the probability that that study will find a statistically significant effect of the stickers at the 95% level if the true effect of the stickers is to decrease the accident rate by 20%.

- For example, we find that there is a 93% chance that the Ugandan RCT will detect a statistically significant effect of the stickers at the 95% level if the true effect of the stickers is to decrease the accident rate by 20%.
- The Ugandan RCT is the most powerful, and Habyarimana & Jack 2011 the least. The differences are driven by differences in the sample size and the accident rate.

1.2.2 Using power calculations to create weights for meta-analysis

- We propose that each study receives a weight proportional to its relative statistical power in order to conduct a meta-analysis.
- We normalise the power of each study relative to the power of the Ugandan RCT (the most powerful), to create those weights.
- We do this for each postulated true treatment effect, and present the results in Table 2.
- We treat the 20% true treatment effect as our preferred assumption.
- **Conclusion: If we only want to adjust for the difference in statistical power across the studies in a meta-analysis, we should apply a weight of 1 to the Ugandan RCT, a weight of 0.95 to Habyarimana & Jack 2015, a weight of 0.34 to the Tanzanian RCT, and a weight of 0.26 to the Habyarimana & Jack 2011.**
- Of course, there may be other reasons for weighting one study more than another, this is only an adjustment for statistical power.

Table 1: Power Calculations

	True Effect Size		
	20%	10%	5%
Tanzania	**0.31**	0.11	0.06
Uganda	**0.93**	0.39	0.13
Habyarimana & Jack (2011)	**0.24**	0.09	0.06
Habyarimana & Jack (2015)	**0.88**	0.34	0.12

Table 2: Relative Weights for Meta-Analysis

	True Effect Size		
	20%	10%	5%
Tanzania	**0.34**	0.29	0.49
Uganda	**1**	1	1
Habyarimana & Jack (2011)	**0.26**	0.24	0.46
Habyarimana & Jack (2015)	**0.95**	0.88	0.92

Notes:

- Statisticians caution against doing retrospective power analysis. The reason is as follows. To conduct power analysis, researchers often use the observed treatment effect as the postulated true treatment effect in their calculations. They then use the calculated power of the study as if it were an additional piece of information over and above the p-value of the test in the study, and use it to modify the interpretation of the study they had made based on the p-value alone. (E.g. it is typically used to make some kind of statement like: 'the null hypothesis was not rejected, but the statistical power of the test was low, so we can't really take this as strong evidence in favour of the null hypothesis'). To do power analysis in this way is incorrect. The reason is that if we use the observed treatment effect as the postulated true treatment effect, there is a 1:1 relationship between the power of the study and the p-value of the test in the study, and so the power of the study provides exactly no additional information over and above the p-value.

- The power analysis that we have done for the two Kenyan RCTs here does not fall into this trap for two reasons:

    1. We do not use the estimated treatment effects from the studies (50% and 25% effects) as the postulated treatment effects for the power calculations. We rather use our own priors about the likely size of the effects.

    2. More importantly, we are not using the power estimates to 'update' our interpretation of the tests done in the study based on the p-values. E.g. the results in the Kenyan RCTs are significant. The wrong way to use the power calculations, falling into the trap above, would be to say: 'Habyarimana & Jack 2011 finds a significant effect even though it had low power. So that study provides

even stronger evidence in favour of the alternative hypothesis.' Rather, we are calculating the relative power of each study as if we had never seen the results of the study, for the purpose of creating meta-analysis weights. In this sense, what we are doing is not retrospective at all, it is really ex ante.

1.3 Power calculations: Notes on Stata code

We use the Stata command **power**, which allows the user to estimate the power of a study for a given sample size. More specifically, we use the Stata command **power twoproportions** throughout:

```
power twoproportions X1 X2, n1(S1) n2(S2) alpha(0.05)
```

Where: X1 is the initial accident rate, X2 is the postulated accident rate for the treatment group based on the assumed true treatment effect size, S1 is the size of the control group sample and S2 is the size of the treatment group sample. For example, for the Tanzanian RCT and a 20% true treatment effect, we use the code:

```
power twoproportions 0.032 0.0256, n1(3197) n2(2789) alpha(0.05)
```

2 Spillovers

There are two spillover effects in Habyarimana & Jack 2011 and Habyarimana & Jack 2015:

- **Spillover 1**. Drivers in control vehicles have previously driven treatment vehicles. If they have internalised previous passenger complaints, they may drive more safely in control vehicles too.
- **Spillover 2**. Passengers in control vehicles may have previously ridden in treatment vehicles. They know about the stickers already, and so may complain in the control vehicles too.

2.1 Adjustment for Spillover 1

Define two parameters:

- s^{driver} - the fraction of control vehicle journeys driven by a driver who has previously driven a treatment vehicle.
- i - the strength of the treatment for a 'treated' driver in a control vehicle, as a fraction of the strength of the treatment for a 'treated' driver in a treatment vehicle. E.g. if $i = 0.2$ then remembering a passenger telling you to drive more safely in a previous journey has one fifth the influence of having a passenger tell you to drive more safely in the current journey.

Then $s^{driver} \times i$ tells us the treatment effect in the control vehicles, as a proportion of the treatment effect in treatment vehicles. Then the treatment effect that we estimate is equal to $(1 - (s^{driver} \times i))$ of the true treatment effect. Therefore, defining β as the true treatment effect, and $\widehat{\beta}$ as the estimated treatment effect:

$$\beta = 1/(1 - (s^{driver} \times i)) \times \widehat{\beta}$$

2.2 Adjustment for Spillover 2

Define two parameters:

- $s^{passenger}$ - the fraction of passengers in control vehicles who have previously been in a treatment vehicle.
- e - the strength of the treatment for a 'treated' passenger in a control vehicle, as a fraction of the strength of the treatment for a 'treated' passenger in a treatment vehicle. E.g. if $e = 0.2$ then remembering a sticker from a previous journey has one fifth the effect of seeing a sticker during the current journey on the probability that the passenger complains.

Then, just as before, $s^{passenger} \times e$ tells us the treatment effect in the control vehicles, as a proportion of the treatment effect in treatment vehicles. Then the treatment effect that we estimate is equal to $(1 - (s^{passenger} \times e))$ of the true treatment effect. Therefore:

$$\beta = 1/(1 - (s^{passenger} \times e)) \times \widehat{\beta}$$

2.3 Plausible assumptions for parameter values, and the overall adjustment

Parameter assumptions:

- s^{driver}. According to Table 1 in Habyarimana & Jack 2011, 72% of drivers in the control vehicles drove only one vehicle. So then 28% of control vehicles are driven by a driver who has driven another vehicle. As an upper bound for the spillover adjustment, assume that all these drivers have driven a treatment vehicle. Then (assuming that all drivers drive the same number of journeys), an upper bound for $s^{driver} = 0.28$. (An equivalent figure is not reported in Habyarimana & Jack 2015).

- $s^{passenger}$. According to Habyarimana & Jack 2015: '...passengers nearly certainly will ride on both treated and untreated vehicles', p.E4662. Therefore, $s^{passenger} = 1$.

- i. We have very little information with which to make an assumption about this. It will depend on things like how often the driver has driven a treatment vehicle in the past, and how much he has internalised previous complaints. We use $i = 0.2$.

- e. We have very little information with which to make an assumption about this. We use $e = 0.2$.

Overall adjustment:

$$\begin{aligned}\beta &= 1/(1 - (s^{driver} \times i)) \times 1/(1 - (s^{passenger} \times e)) \times \widehat{\beta} \\ &= (1/0.944) \times (1/0.8) \times \widehat{\beta} \\ &= 1.324 \times \widehat{\beta}\end{aligned}$$

3 Sources

- Habyarimana & Jack 2011

- Habyarimana & Jack 2015