Annex III. Power analysis to estimate the sample size

In this project, we test the null hypothesis that there is no significant difference in the proportion of agricultural area under residue burning in villages with Happy Seeder adoption and villages without. The aim of the power and sample-size (PSS) analysis is to determine the number of villages to be sampled, where village surveys and key informant interviews will be conducted and for which the remote sensing data will be generated. By carrying out the PSS analysis, we can estimate the sample size required to achieve the desired power of a test in the planned study (Djimeu and Houndolo 2016). For the village-level impact assessment, we can divide the sample villages into two groups – villages where Happy Seeder technology is prevalent, and villages without the technology (or where its adoption is marginal with <5% wheat area under the Happy Seeder). For the PSS analysis, we derive information from the CIMMYT 2018 survey. Analysis of the data indicated that –

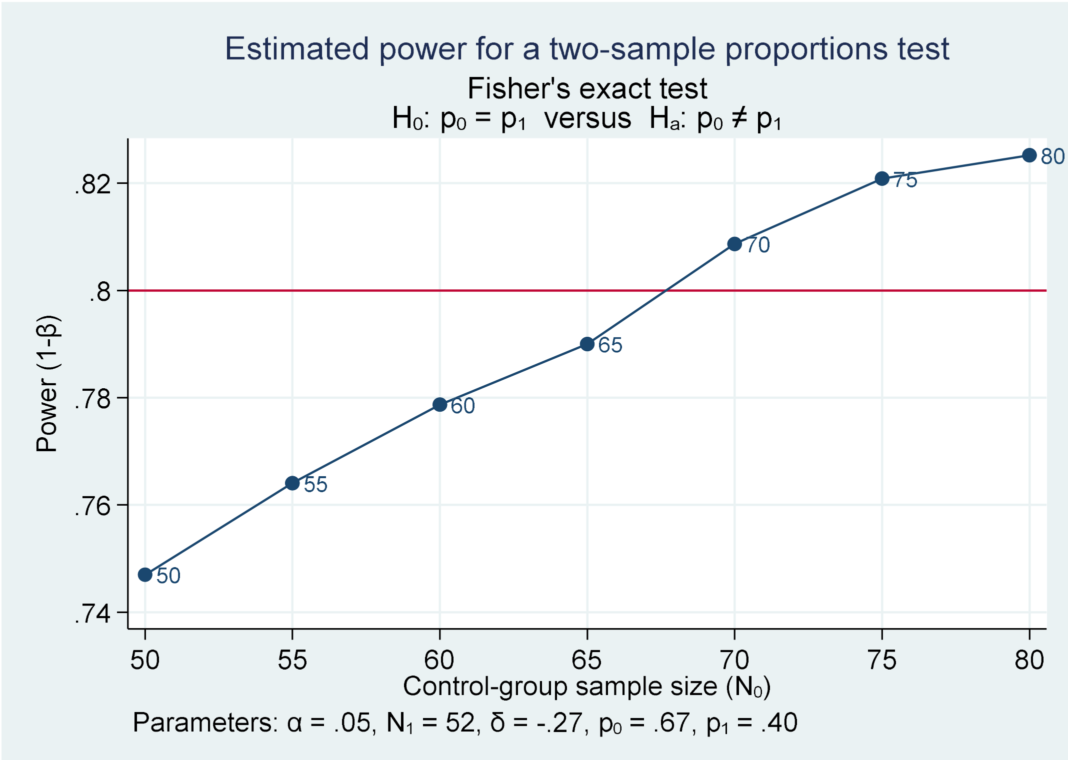
(a) In about 67% rice-wheat area under the traditional tillage systems, residue burning is practiced,
(b) With the Happy Seeder adoption, residue burning is limited to 5% area ($\mathrm{B}_1 = 0.05$), and
(c) In villages with the Happy Seeder technology, 43% wheat area is under it (A).

We assume that in villages without the Happy Seeder technology, the rate of residue burning would be comparable to that of the non-adopter group of villages with the technology.[1] If so, residue burning would be prevalent in 67% of wheat area in 'control' villages (p_0). Comparable estimates were obtained from satellite imageries (Chawala and Sandhu 2020). The rate of residue burning in villages with the average rate of adoption of Happy Seeder technology would be $\mathrm{p}_1 = [\mathrm{B}_1 \times A] + [\mathrm{p}_0 \times (1 - A)] = 0.40$. With control-group proportion $\mathrm{p}_0 = 0.67$, treatment-group proportion $\mathrm{p}_1 = 0.40$, default power of 0.80, and significance level $\alpha = 0.05$,[2] we estimate the sample size as 106. This would be ideal to test the hypothesis $H_0\colon \mathrm{p}_0 = \mathrm{p}_1$ versus $H_a\colon \mathrm{p}_0 \neq \mathrm{p}_1$.

The CIMMYT 2018 survey covered about 52 villages where Happy Seeder technology is present, where re-survey is necessary to estimate the continuity in technology adoption process and changes in the institutional environment. We hence estimate the number of control-group of villages (without or marginal adoption of the Happy Seeder) using the PSS calculation. We found that by adding 70 control villages to the 2018 CIMMYT sample, the sampling process would yield a power above 0.80 (Figure 1). In total, the minimum number of villages to be included in the village survey would be 122.

[1] There could be some technology spillovers in villages with the Happy Seeder technology, which we cannot account for in the PSS calculation due to data limitation.
[2] We used Stata 16.0 command '*power twoproportions*' for estimating the sample size.

Figure 1.

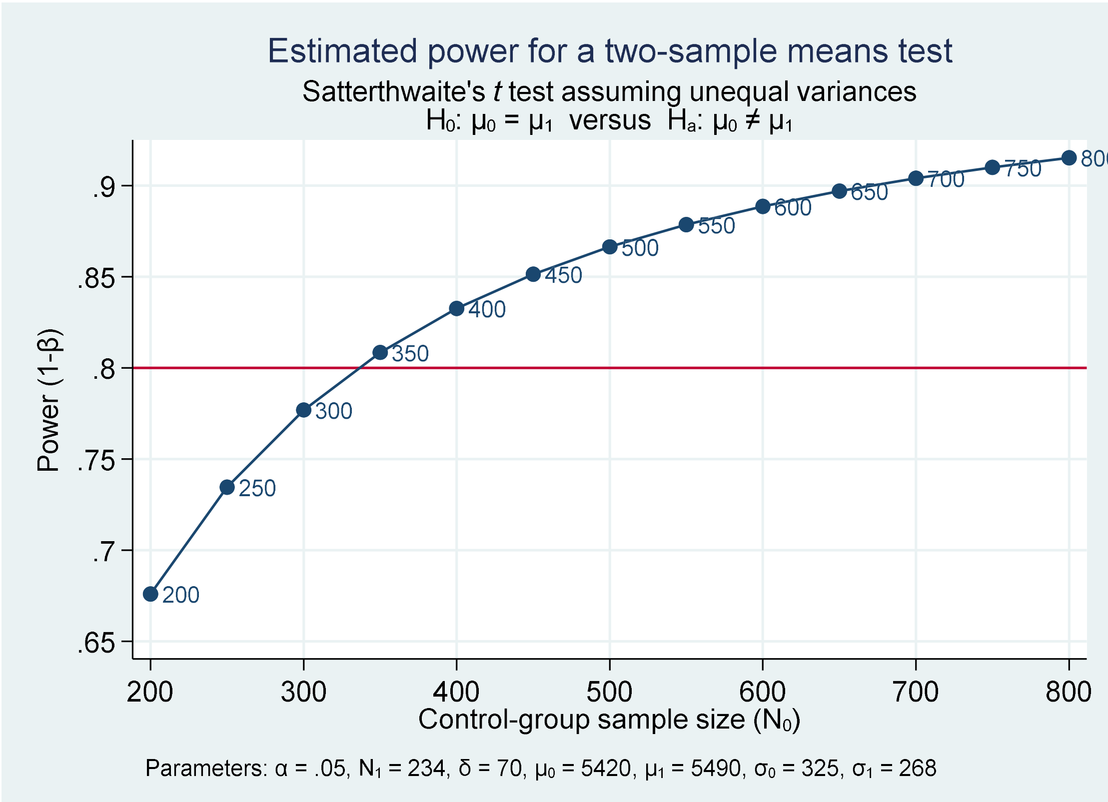


In case of household survey, due to the huge difference in proportion of area under residue burning between adopting households of the Happy Seeder technology and non-adopting households (i.e., 0.67 versus 0.05),[3] a small sample size would be sufficient to achieve the desired power of a test. However, the household survey data are required also for assessing the yield impact of the technology. A quick analysis of 2018 CIMMYT survey indicated that the adopters of Happy Seeders get 5490 kg wheat grain per hectare (with std. deviation of 268), and non-adopters get 5420 kg wheat grain per hectare (with std. deviation of 328). There were 234 technology adopters in 2018 survey. Assuming no huge change in the adoption rate of Happy Seeders between 2018 and 2021, the minimum number of non-adopters required for 234 adopters would be about 350, with default power of 0.80 and significance level $\alpha = 0.05$ (Figure 2).[4] The 2018 CIMMYT survey included 561 non-adopters, and hence we decided to retain the sampling frame, and revisit the same farmers in 2021. The panel dataset would also help identify the temporal changes in the diffusion process (e.g., dis-adoption of technology).

---

[3] Adopters use the Happy Seeder technology in 95% of their cultivated area.
[4] We used Stata 16.0 command '*power twomeans*' for estimating the sample size.

Figure 2.



References

Chawala, Pratika; Sandhu, H. A. S. (2020): Stubble burn area estimation and its impact on ambient air quality of Patiala & Ludhiana district, Punjab, India. In *Heliyon* 6 (1), e03095. DOI: 10.1016/j.heliyon.2019.e03095.

Djimeu, Eric W.; Houndolo, Deo-Gracias (2016): Power calculation for causal inference in social science: Sample size and minimum detectable effect determination. In *Journal of Development Effectiveness* 8 (4), pp. 508–527. DOI: 10.1080/19439342.2016.1244555.