

(Very!) Preliminary Report - ColaLife Project

1 Methods

The following assumes to have already introduced the context and the data...

1.1 Analyses on the Aggregated Data

The present study aims to investigate whether the co-pack introduction has significantly increased the probability with which an under-five child affected by diarrhoea is given both ORS and zinc as treatment. The available data allows to estimate the probability with which an under-five child is treated correctly from diarrhoea as the proportion of correctly-treated children (CTC) within the sample. The true proportion of CTC (to be imagined, for example, as the one computed across all Zambian health centres, for most of which data is not available) may however be different from the one observed in the available sample. For this reason, we associate a 95% confidence interval (CI) to the previous estimate, using the Clopper-Pearson method for an exact binomial CI. The interval is expected to contain the true proportion of CTC with 95% confidence.

The above sample estimate and associated CI concern the proportion of CTC in a specific context, typically either before or after co-packaging introduction. Of particular interest to this work, however, is to test whether a significant increase in such proportion has taken place upon introduction of the ORS and zinc co-pack. To this aim, we perform a one-sided test of homogeneity of binomial proportions, testing the null hypothesis that the proportions of CTC before co-pack introduction and after co-pack introduction are equal to each other, against the alternative that the latter is higher than the former. We perform this and all other tests of this work at the 5% significance level: that is, a

p -value < 0.05 is considered as evidence that a significant increase in the proportion of CTC¹ has taken place after co-packaging introduction.

The above choice of significance level is the most commonly adopted choice across the applied statistics literature, both in medicine and beyond. However, we wish to stress that the 5% choice is purely a convention. Two p -values such as, *e.g.*, $p = 0.045$ and $p < 0.0001$, are both considered significant at the 5% level. However, the level of evidence that they carry against the null hypothesis (and in favour of the alternative) is very different in the two cases. A p -value < 0.0001 reveals that, if the null hypothesis were true, the observed data would have a probability of happening lower to one in ten thousand cases. This clearly represents a much stronger evidence against the null hypothesis than if the data could happen in about one every twenty cases ($p \approx 0.05$) under the null hypothesis.²

The p -value of the above homogeneity of proportions test is a measure of statistical significance. As such, it does provide a quantitative measure of the effect size of co-packaging on the proportion of CTC. To quantify this, we employ two measures: the difference between the two proportions and the rate ratio.

The difference between the two proportions is estimated via the sample difference and the 95% Wald CI [1]. The interval is constructed using the large-sample normal approximation and provides a range of values that are expected to contain the true difference of the two proportions, with 95% confidence. The rate ratio is instead computed as the ratio between the proportions of CTC after co-pack introduction, and the proportion before co-pack introduction. It quantifies how many times more (or less) likely a child is to be correctly treated from diarrhoea after co-pack, than they were before. A CI for the rate ratio is computed following [1, 2].

¹I guess we could introduce an acronym for the “proportion of correctly treated children”, since this is the main quantity the whole work focuses on.

²We can discuss whether similar considerations/explanations should be included. I feel they can have value since people may not be really familiar with statistical tests or interpretation of p -values, but we’ll discuss this together with all the rest in a call.

2 Results

2.1 Agglomerated Results, over all seven centres

The data (to be summarised in a table..)

Before (the) co-pack introduction, a total of 176 cases of diarrhoea were recorded across the seven centres, of whom 77 were given both ORS and zinc (11 of these cases where given less than ten zinc tablets though).

After the co-pack introduction, 389 total cases were instead recorded across the seven centres, of whom 337 correctly treated (all of whom with at least ten zinc tablets). This leads to the following results.

Before co-pack, we obtain a sample proportion of CTC equal to 0.44, CI: (0.36, 0.51). After co-pack instead, the estimated proportion of CTC rises to 0.87, CI: (0.83, 0.90). *Note that the “before” estimate is computed by classifying as correctly treated those children who were given less than 10 zinc tablets (otherwise, the estimate becomes 0.38 with a CI of (0.30, 0.45)).*

A one-sided test of homogeneity of proportions is highly significant ($p < 0.0001$), suggesting that the data carries strong evidence towards the hypothesis that the proportion of CTC is higher after co-pack introduction. The 95% CI for the difference between the two proportions is (0.35, 0.51), with a central estimate of 0.43. Such difference is a number between -1 and 1 : positive (negative) values signify that the proportion of CTC has increased (decreased) after the co-packaging introduction. Figure ** shows the CIs of the two proportions and of their difference, on the respective scales.

The 95% CI for the rate ratio of the two proportions equals (1.68, 2.37). In other words, with high confidence, we can state that the rate at which a child is correctly treated from diarrhoea has become between 68% and 137% higher after the introduction of the co-pack, than it was before. The 99% CI is only slightly wider, (1.61, 2.52).

References

- [1] Alan Agresti. *An Introduction to Categorical Data Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, second edition, 2002.
- [2] Jun-Mo Nam. Confidence limits for the ratio of two binomial proportions based on likelihood scores: non-iterative method. *Biometrical Journal*, 37(3):375–379, 1995.