**Explanatory note for dataset used to assess relationship beteween millet cultivation and malnutrition patterns in India at district level**

For the analysis, district level dataset was prepared from the three data sources

1) Government of India census data (2011). Available from: <https://censusindia.gov.in/>

2) National Family Health Survey 4 (NFHS4) data was downloaded from DHS website: <http://dhsprogram.com/>

3) For the cultivation data, we used DACNET, a web-based land use statistics information system maintained by the Agriculture Informatics Division of the National Informatics Centre of the Government of India <https://aps.dac.gov.in/LUS/Index.htm>

From each of the three datasets, the following data were extracted

1. From the 2011 census data, district-wise total population
2. From the NFHS4 data,
   1. using an appropriate weights BMI of women 15-49 years age group and short stature was calculated from women’s dataset
   2. children wasting and stunting was calculated from the children’s dataset
   3. Proportion of people in wealth quintiles was calculated from household dataset
3. Various crop data is available in state-wise reports compiled by the Ministry of Agriculture and Farmers Welfare. We extracted district-level area under cultivation of cereals: rice, wheat, maize, ragi and millets (by type as defined above) into a spreadsheet. Data was from the latest state-wise reports available at the time of analysis at DACNET (data for most states ranged for years between 2014-17 except Maharashtra 2002-03, Manipur 2004-05 and Gujarat 2007-08; all data in hectares converted to acres).

**Creation of master data set**

Using district names as the common variable in all three datasets, they were merged. Any errors due to district spellings and duplicate district names across differing states were handled with caution to ensure proper merging.

For each district we estimated the population of poor by multiplying the census figures for population of the district by the proportion of the population in the fourth and fifth wealth quintiles (from NFHS4). This was based on the assumption that subsistence millet consumption is largely restricted to poor small land-holding farmers45. We thus calculated the District Subsistence Cultivation Quantum(DSCQ) for each cereal by multiplying the per-capita area under cultivation of cereal with the population of poor as estimated above. We formulated this index as it appeared best suited to represent the area of subsistence farming in a district with a direct bearing on prevalence (or lack of) malnutrition in that region

**Calculation of a normalized district subsistence cultivation quantum (DSCQ)**

Step 1: All the millet data was converted to acre from hectare by multiplying 2.47105 to millet hectare data.

Step 2: (Millet cultivated/Total population) \* Proportion of Poor

Step 3: Millets data was normalised using logarithmic transformation