**The Data**

To measure the effect of CHILDREN´s engagement on the organizations we use the data they collected from 2012 to 2020. In each year they send a survey to the organizations with several questions about the previous year. The number of organizations varies among the years and increases over time, from 52 in the 2012 survey to 73 in the survey from 2020. In some organizations one employee fills in the survey and in others they do it as a team. Since the children and adolescents are not questioned directly, all responses are documented through the perception of the employees. The number of variables varies over time as well. Included are numbers like the average eaters per meal or the amount of money they provide to the organizations but also general questions. For instance, CHILDREN asks the average amount of kids with a better confidence or an improved dietary knowledge in the specific organization. This part of the survey must be answered on a scale from zero (no kids) to four (all kids). If an organization do not answer a question, this is documented as a “99”. We worked with the statistical program “R” and therefore changed the format from 99s to NA´s (not available) to avoid distortions. The surveyed variables change over the years, but some of them are included every year.

However, we did several steps to get a full dataset we could work with. The data was divided into one dataset for each survey from 2012 to 2020, but we only use the surveys till 2019 since in 2020 some organization-ID´s occurred several times and the data for 2020 were incomplete. Since each survey includes data about the year before, we changed the names of the dataset to the corresponding year and finally used the years from 2011 to 2018. Moreover, we outlined a hierarchical file structure, enabling us to use relative file paths throughout. This makes a quick work with R possible since we only use paths relative to the working directory. Afterwards we made sure that variables with names containing non-standard characters like German “Umlaute” are correctly read in and established naming conventions. We created a file reading the excel sheets and we reviewed and aligned new English-language variable names across the years. Moreover, we systematically compared variable names between years by creating a correspondence table, ordered first by variables of 2019, then of 2018 and so on. To ensure the comparability between the years, we gave all variables from the different years that equal each other the same name. As a next step, we merged the different datasets to one dataset, including all years and variables CHILDREN collected. For an efficient and clear data structure, we created a function that automatically changed the data type of all variables from "character" to "ordinal" and added several versions for each initially metric encoded variable afterwards. The three variants are ordinal, standardized and weighted (FUßNOTE). Furthermore, we created more new variables: We used the information CHILDREN gave us in another excel-sheet to assign the German states to the corresponding organization-ID and created dummy-variables for each ID, every year and a treatment dummy that will be explained in a later section.

The final dataset we worked with is structured as follows: Each row represents one organization-ID with the answers the organization gave in the specific year. The questions are divided in two categories: the variables regarding to the Mittagstisch, answered by all organizations since they are all part of this program and the Entdeckerfonds variables, answered by the organizations that take part on the Entdeckerfonds program in the respective year. Including the years from 2011 to 2018 and all variables we created, the final dataset has X observations of Y variables.