**Caregiver survey data cleaning plan**

1. **Progress tracking** 
   1. Check progress by camp; check number of surveys completed; check number of surveys remaining
      1. Columns: Camp ID, surveys needed, number of surveys (yes consent) completed, number of surveys remaining, points given, number of points ineligible for survey, number of points where consent not given
2. **GPS**
   1. Identify any points recorded that day that fall outside of 25m from the assigned GPS point
      1. Columns: Date, enumerator ID, UUID, Camp name
3. **Survey duration using audit function**
   1. Identify any surveys completed in less than 30 minutes
      1. Columns: Date, enumerator ID, UUID, survey duration in minutes (only if less than 30 minutes)
   2. Calculate average time per survey across all enumerators
      1. Columns: Number of surveys completed, mean time duration, median time duration
   3. Calculate average time per survey per enumerator
      1. Columns: Enumerator ID, number of surveys completed; mean time duration; median time duration; percent difference from average mean time
4. **Enumerator** 
   1. Identify if enumerators record “don’t know/prefer not to answer” for over half of the questions across the following sections: HarmThreats, HarmWork, HarmMarr, HarmVio, HarmNegl, HarmWork2, HarmMarr2, HarmInfo)
      1. Columns: Enumerator ID, number of surveys in which over half of answers are “don’t know/prefer not to answer”
5. **General checks**
   1. Identify any duplicate UUIDs
      1. Columns: Date, enumerator ID, UUID (if duplicate)
   2. Identify any surveys that do not end on the same day they started; any surveys that end before they start; any surveys made before the first day of data collection (pilot: 09 Feb); any surveys made in the future
      1. Columns: Date, enumerator ID, UUID, survey start time, survey end time
6. **Consent**
   1. Identify any surveys in which time spent on consent is less than 45 seconds (consent\_message: Hello my name is \_\_\_\_\_\_. I work for REACH. Together with UNICEF, we are currently conducting a survey…)
      1. Columns: Date, enumerator ID, UUID, time spent on consent (if less than 45 seconds)
7. **HarmThreats**
   1. Identify any surveys in which time spent on threats question (threats\_30: In the past 30 days, which threats have you been concerned about for your child's safety?) is less than 45 seconds
      1. Columns: Date, enumerator ID, UUID, amount of time on question (if less than 45 seconds)
8. **Agree/disagree questions (****HarmWork, HarmMarr, HarmVio, HarmNegl)** 
   1. Identify any enumerators that have a clear pattern of responses to the agree/disagree questions across their surveys (mostly agree, mostly disagree, exactly half agree/half disagree)
      1. Columns: Enumerator ID, UUID, number of agrees (if greater than 24), number of disagrees (if greater than 24), number of agrees = number of disagrees
9. **“Other” responses**
   1. Identify all “other” responses by question
      1. Columns: Date; enumerator ID; other responses