**Data cleaning plan – J-MSNAs 2021**

For all checks, please provide the following columns:

* Respondent ID
* Enumerator ID, enumerator name, TL name
* Uuid
* Survey loop (HH, ind1, ind2)
* instance name (including parent and repeat instance name, where relevant)

With the exception of the progress tracking, checks should only be run across surveys where informed\_consent == “yes”

CAMPS ONLY

HC ONLY

|  |  |  |
| --- | --- | --- |
| **1. Daily progress tracking** | **Priority** | **Comment** |
| Track completed surveys by camp/union   * Total number of surveys per camp/union by response option for “reached” * Total number of surveys per camp/union by response option for “informed\_consent” | 1 | Will allow us to track phone coverage and progress |
| **2. Survey verification** |  |  |
| * Duplicate respondent IDs | 1 |  |
| **3. Enumerator and survey duration issues** |  |  |
| Surveys completed in less than 25 minutes | 1 |  |
| Duration statistics by enumerator – average time per survey per enumerator  Duration by group | 1 |  |
| Number of surveys completed per enumerator per day | 1 |  |
| Check if enumerators selected dont\_know for more than 15 questions (to be adjusted later on) | 1 |  |
| Number of issues flagged by enumerator not clear | 1 |  |
| **4. Logical inconsistencies** |  |  |
| **clinics**: flag if more than 3 selected – add column showing the selected options | 2 |  |
| **individuals\_difficulties**: flag if more than 33% of HH members selected – add column indicating number of HH members selected, hh\_size, and the values for indv\_difficulty\_seeing, indv\_difficulty\_hearing, indv\_difficulty\_walking, indv\_difficulty\_remembering, indv\_difficulty\_selfcare, indv\_difficulty\_communicating for each selected individual | 1 | Should flag unusually high number of “persons with disabilities” in HH |
| **shelter\_issues\_reason**: flag if more than 3 selected – add column showing selected options | 2 |  |
| **shelter\_issues\_reason**: flag any of the following combinations – add columns showing values for shelter\_issues\_reason, leaks\_during\_rain, limited\_ventilation, dirt\_debris, lack\_of\_insulation, collapse\_living\_there   * == “damage\_roof” & leaks\_during\_rain != “yes” & lack\_of\_insulation != “yes” * == “damage\_windows” & leaks\_during\_rain != “yes” & lack\_of\_insulation != “yes” * == “structure\_unstable” & collapse\_living\_there != “yes” | 1 | Should flag if a reason is reported that does not seem to be related to the reported enclosure issue – not necessarily an issue, i.e. can probably often be explained by the enumerator but good to check |
| **Improvement**: flag if more than 5 selected – add column showing all selected options | 2 |  |
| **improvement\_reason**: flag if did\_not\_receive\_any\_shelter\_support & sold\_materials selected – add column showing selected options | 1 | Should flag if the HH reports not having received materials and having sold received materials |
| **Improvement\_reason**: add columns showing improvement\_reason, leaks\_during\_rain, limited\_ventilation, dirt\_debris, lack\_of\_insulation, collapse\_living\_there   * flag if no\_able\_bodied\_person & at least one male person in the household aged 18-59 that is not selected under individuals\_difficulties * flag if no\_need\_to\_improve & leaks\_during\_rain !=”yes” & limited\_ventilation !=”yes” & dirt\_debris !=”yes” & lack\_of\_insulation !=”yes” & collapse\_living\_there !=”yes” | 1 | Should flag if no able-bodied person is reported but there is a not disabled adult male in HH |
| **shelter\_materials\_source**: flag if more than 3 selected – add column showing selected options | 2 |  |
| **shelter\_paid**: flag if more than 2 selected – add column showing selected options | 2 |  |
| **group\_nfi**: flag if more than 8 are NOT “yes” – add column showing all that are NOT “yes” | 1 |  |
| **alternative\_cooking\_fuel**: flag if more than 3 selected – add column showing selected options | 2 |  |
| **health\_distance**: flag if >90 and flag if +/-2sd from mean – add columns showing values for health\_distance, and health\_transportation | 1 | Should flag unusually large or small values |
| **health\_barriers\_not\_accessed, health\_barriers\_accessed, health\_barriers\_no\_need**: flag any of the following combinations (for either of the 3 – only one those three questions will be asked to every HH) – add columns showing selected values for those 3 barrier questions (one column is enough, since only 1 of those questions would be asked at a time), health\_distance, health\_transportation, paid\_healthcare:   * == (“no\_functional\_facility\_nearby” | “health\_services\_are\_too\_far\_away”) & health\_distance < 30 * == “cannot\_afford” & paid\_healthcare == “yes” | 1 | Should flag if travel time was short but health facility is reported to be far  Should flag if HH paid for health care and says it couldn’t afford health care |
| **group\_enrolment & group\_distance\_learning**: flag if for any gender/age group the number of individuals in distance learning is larger than the number of individuals enrolled, e.g. if distance\_learning\_girls\_3 > enrolment\_girls\_3 – add columns showing the distance learning/enrolment values for each age/gender group | 2 | Should flag if number of individuals in an age/gender group that followed some distance learning is larger than the number of individuals in that group that was enrolled pre-COVID |
| **group\_formal\_enrolment, group\_informal\_enrolment & group\_distance\_learning**: flag if for any gender/age group the number of individuals in distance learning is larger than the number of individuals enrolled in either formal or informal learning, e.g. if distance\_learning\_girls\_3 > sum(formal\_enrolment\_girls\_3, informal\_enrolment\_girls\_3) – add columns showing the distance learning/enrolment values for each age/gender group | 2 | Should flag if number of individuals in an age/gender group that followed some distance learning is larger than the number of individuals in that group that was enrolled pre-COVID |
| **group\_send\_back**: flag if for any gender/age group the number of individuals that will be sent back/has returned is larger than the number of individuals enrolled/in distance learning, e.g. if send\_back\_girls\_3 > (distance\_learning\_girls\_3 | enrolment\_girls\_3) – add columns as above | 2 | Should flag if for any group more individuals are being sent back/have returned than were enrolled previously or did distance learning |
| **group\_send\_back**: flag if for any gender/age group the number of individuals that will be sent back/has returned is larger than the number of individuals enrolled in any form of learning/in distance learning, e.g. if send\_back\_girls\_3 > (distance\_learning\_girls\_3 | sum(formal\_enrolment\_girls\_3, informal\_enrolment\_girls)) – add columns as above | 2 | Should flag if for any group more individuals are being sent back/have returned than were enrolled previously or did distance learning |
| **distance\_learning\_barriers\_girls/distance\_learning\_barriers\_boys** (2 separate checks): for both, flag any of the following combinations – add columns showing values for distance\_learning\_barriers\_girls/distance\_learning\_barriers\_boys, the “count” fields added below, solar\_lamps, living\_space\_issues, the “married” fields added below, child\_marriage:   * == “working\_outside\_home” & child\_working\_long\_hours != “yes” * == “lack\_content\_older\_children” & girl\_15\_18\_count == 0 & girl\_19\_24\_count == 0 (or for boys) * == “lack\_content\_younger\_children” & girl\_3\_count == 0 & girl\_4\_5\_count == 0 (or for boys) * == “lack\_light” & (solar\_lamps == “yes” | living\_space\_issues != “lack\_lighting\_inside”) * == “marriage\_pregnancy” & (girls\_married == 0 | child\_marriage != “yes”) (or for boys) | 1 | Should flag if barrier to distance learning was that the child was working but this is not indicated as livelihood-based coping mechanism – not necessarily an issue, since questions are not exactly the same, but good to check  Should flag if lack of content for older children indicated as barrier but no children aged 15-24 in HH  Should flag if lack of content for younger children indicated as barrier but no children aged 3-5 in HH  Should flag if lack of light indicated as barrier but HH has solar lamps or did not report lack of light under living space issues  Should flag if marriage/pregnancy indicated as barrier but no married children reported |
| The same checks as above for: **send\_back\_barriers\_girls, send\_back\_barriers\_boys, send\_back\_challenges\_girls, send\_back\_challenges\_boys, return\_barriers\_girls, return\_barriers\_boys** | 1 | As above |
| **nutrition\_barriers\_not\_accessed**: flag if child\_taken\_no\_support\_HH\_count > 0 and NONE of the following selected – add columns showing selected values for barrier question, child\_taken\_no\_support\_HH\_count, child\_not\_taken\_after\_referral\_HH\_count   * already\_referred * not\_admitted * waiting\_times * language\_barriers | 1 | Should flag if there was a child taken to nutrition facility but did not receive support but as barrier none of the indicated options is given |
| flag if (nutrition\_messaging == “yes” & plw\_nutrition\_messaging == “no”) | (nutrition\_messaging == “no” & plw\_nutrition\_messaging == “yes”) – add columns with values for those fields | 2 | Should flag if the same message was received for children but not PLW or vice versa |
| flag if plw\_admitted > plw\_sfs – add columns with values for those fields | 1 | Should flag if more PLW were admitted than received support |
| flag if (plw\_iron\_folic\_acid > 0 & adol\_girls\_folic\_acid\_tablets == 0) | (plw\_iron\_folic\_acid == 0 & adol\_girls\_folic\_acid\_tablets > 0) – add columns with values for those fields | 2 | Should flag if PLW received tablets while adolescent girls didn’t or vice versa |
| **food\_barriers**: flag any of the following combinations – add columns showing values for food\_barriers, feedback:   * == “no\_assistance” * == “feedback” & feedback != “did\_not\_know\_where” & feedback != “could\_not\_go\_out” * == “lack\_response” & feedback != “did\_but\_it\_was\_not\_responded” & feedback != “did\_but\_the\_response\_was\_not\_satisfactory” | 1 | Should flag if HH didn’t receive any assistance  Should flag if problems with feedback related to food assistance indicated but not under CwC |
| **group\_livelihood\_coping**: flag any of the following combinations – add columns showing values for begging, depending\_on\_food\_rations, selling\_food\_rations, selling\_nfis, income\_source:   * begging == “yes” * depending\_on\_food\_rations == “yes” & (income\_source == “salaried\_work” | income\_source == “casual\_labor” | income\_source == “business” | income\_source == “cash\_for\_work” | income\_source == “donations”)) * depending\_on\_food\_rations == “yes” & (income\_source == “salaried\_work” | income\_source == “casual\_labor” | income\_source == “business” | income\_source == “government\_benefits” | income\_source == “cash\_for\_work” | income\_source == “donations”)) * (selling\_food\_rations == “yes” | selling\_nfis == “yes”) & income\_source != “sale\_assistance” * child\_working\_long\_hours == “yes” AND NO individual under the age of 18 indicated to be working under individuals\_income * child\_marriage AND (girls\_married == 0 AND boys\_married == 0) * risk\_jobs == “yes” AND NO individual indicated to be working under individuals\_income * adults\_working\_long\_hours == “yes” AND NO individual aged of 18 and above indicated to be working under individuals\_income | 1 | Should flag if depending on food rations/support as only food/income source indicated as coping mechanism but other income sources are indicated  Should flag if selling rations/NFIs indicated as coping mechanism but not as income source  Should flag if children/adults indicated as coping strategy but not source of income  Should flag if child marriage indicated here but no married children in HH |
| **coping\_reason**: flag if more than 5 selected – add column showing selected options | 2 |  |
| **coping\_reason**: flag any of the following combinations – add columns for coping\_reason, food\_expenditure, health\_expenditure paid\_healthcare, all the enrolment/distance learning/send back fields, education\_expenditure, infreq\_nfi\_expenditure, shelter\_paid, rent\_expenditure, shelter\_expenditure, improvement, fuel, fuel\_expenditure, freq\_nfi\_expenditure, water\_coping, water\_expenditure:   * == “food” & food\_expenditure == 0 * == “healthcare” & (paid\_healthcare != “yes” | health\_expenditure == 0) * == “education” & no children in distance learning (or returned, if relevant), i.e. all age/gender groups for the following are 0: distance\_learning\_girls\_3, send\_back\_girls\_3, etc. * == “education” & education\_expenditure == 0 * == (“clothing” | “hh\_items”) & infreq\_nfi\_expenditure == 0 * == “rent” & (shelter\_paid == “no\_need” | shelter\_paid == “dont\_know” | shelter\_paid == ”other” | rent\_expenditure == 0) * == “shelter” & (improvement == “no\_improvement” | improvement == “dont\_know” | improvement == “other” | shelter\_expenditure == 0) * == “cooking\_fuel” & fuel != “buying\_lpg\_refills” & fuel != “firewood\_purchased” * == “cooking\_fuel” & fuel\_expenditure == 0 * == “hygiene\_items” & freq\_nfi\_expenditure == 0 * == “water” & (water\_coping != “spend\_money” & water\_coping != “buy\_water”) * == “water” & water\_expenditure == 0 | 1 | Should flag if accessing food indicated as reason but no food expenditure  Should flag if accessing health care indicated as reason but not indicated that they paid for health care or no health care expenditure  Should flag if accessing education indicated as reason but no children in distance learning or returned  Should flag if accessing education indicated as reason but no education expenditure  Should flag if clothing or HH items indicated as reason but no NFI expenditure related to clothing/HH items  Should flag if rent indicated as reason but no rent payments indicated  Should flag if shelter repair/maintenance indicated as reason but no improvements made or no repair expenditure  Should flag if fuel indicated as reason but no fuel expenditure  Should flag if hygiene items indicated as reason but no NFI expenditures related to hygiene items  Should flag if water indicated as reason but not buying water not indicated as coping mechanism  Should flag if water indicated as reason but no water expenditure |
| **Income\_source**: flag if more than 5 selected – add column showing selected values | 2 |  |
| **salaried\_work\_amount, casual\_labor\_amount, casual\_labor\_days, business\_amount, business\_days, government\_benefits\_amount, remittances\_amount, family\_friends\_amount, cfw\_amount, cfw\_days, donations\_amount, sale\_assistance\_amount, other\_amount** – for each AND for the sum of all, flag if value more than 2sd +/- mean (add columns showing values, mean, and mean +/- 2sd) | 1 | Should flag unusually high/low values for each and the sum |
| **food\_expenditure, rent\_expenditure, water\_expenditure, freq\_nfi\_expenditure, fuel\_expenditure, transportation\_expenditure** – for each AND for the sum of all, flag if value more than 2sd +/- mean (add columns showing values, mean, and mean +/- 2sd) | 1 | Should flag unusually high/low values for each and the sum |
| **rent\_expenditure:** flag if any of the following combinations – also add columns showing values for shelter\_paid   * Rent\_expenditure > 0 & shelter\_paid == “no\_need” * Rent\_expenditure == 0 & shelter\_paid != “no\_need” | 1 | Should flag if rent expenditure indicated but no rent payments in SNFI section and vice versa |
| **fuel\_expenditure:** flag if any of the following combinations – also add columns showing values for fuel   * fuel\_expenditure > 0 & fuel != “buying\_lpg\_refills” & fuel != “firewood\_purchased” * fuel\_expenditure == 0 & (fuel == “buying\_lpg\_refills” | fuel == “firewood\_purchased”) | 1 | Should flag if fuel expenditures indicated but not buying fuel/firewood in SNFI section and vice versa |
| **shelter\_expenditure, infreq\_nfi\_expenditure, health\_expenditure, education\_expenditure, debt\_expenditure** - for each AND for the sum of all, flag if value more than 2sd +/- mean (add columns showing values, mean, and mean +/- 2sd) | 1 | Should flag unusually high/low values for each and the sum |
| **health expenditure:** flag if any of the following combinations – also add columns showing values for paid\_healthcare   * health\_expenditure > 0 & paid\_healthcare == “no” * health\_expenditure == 0 & paid\_healthcare == “yes” | 1 | Should flag if health expenditure indicated but not paying for health care and vice versa |
| **shelter expenditure:** flag if any of the following combinations – also add columns showing values for shelter\_materials\_source   * shelter expenditure > 0 & shelter\_materials\_source != “purchased” * shelter\_expenditure == 0 & shelter\_materials\_source == “purchased” | 1 | Should flag if shelter expenditure indicated but did not buy shelter materials and vice versa |
| **sanitation\_coping**: flag any of the following combinations – add columns showing values for sanitation\_coping, sanitation\_barriers\_female, sanitation\_barriers\_male:   * == “no\_issue” & sanitation\_barriers\_female != “no\_problem” & sanitation\_barriers\_male != “no\_problem” * != “no\_issue” & sanitation\_barriers\_female == “no\_problem” (or NA, if question was skipped) & sanitation\_barriers\_male == “no\_problem” (or NA, if question was skipped) | 1 | Should flag if no sanitation issues indicated under coping but under barriers and vice versa |
| **child\_needs**: flag any of the following combinations – add columns showing values for child\_needs, leaks\_during\_rain, limited\_ventilation, dirt\_debris, lack\_of\_insulation, collapse\_living\_there, collapse\_with\_other\_hh, with\_other\_hh, collapse\_in\_open, shelter\_type, living\_space\_issues, food\_barriers, needed\_healthcare and clinics for all individuals below the age of 18:   * == “food” & food\_barriers == “no\_challenge” * == “shelter” & leaks\_during\_rain != “yes” & limited\_ventilation != “yes” & dirt\_debris != “yes” & lack\_of\_insulation != “yes” & collapse\_living\_there != “yes” & collapse\_with\_other\_hh != “yes” & with\_other\_hh != “yes” & collapse\_in\_open != “yes” * == “shelter” & leaks\_during\_rain != “yes” & limited\_ventilation != “yes” & dirt\_debris != “yes” & lack\_of\_insulation != “yes” & collapse\_living\_there != “yes” & shelter\_type != “jhuprie” & shelter\_type != “kutcha” & shelter\_type != “none\_with\_other\_hh” & shelter\_type != “none\_in\_open” * == “healthcare” & NO individual below the age of 18 with clinics being NA or ANY of the following: ngo\_clinic, government\_clinic, private clinic | 2 | Should flag if food needs of children not met but no challenges with food assistance reported  Should flag if shelter needs of children not met but no shelter issues reported under Shelter  Should flag if health care needs of children not met but no unmet health care need of a child reported, i.e. either no child needed health care or all children that needed health care sought it at a clinic |
| **5. Other responses** |  |  |
| OTHER answers **csv**   * Add question name and value | 1 |  |

Issues

* Duration file corrupted
* Dont\_know question missing
* Couldn’t find col shelter\_expenditure
* Couldn’t find col health\_expenditure
* Couldn’t find col shelter\_paid
* Couldn’t find col education\_expenditure