

Codebook: Returnee Integration to South Sudan and Educational Attainment for Girls

Erica Criollo

2025-04-21

Overview of Data

The Sudan Crisis – Returnee Reintegration & Perceptions Survey was conducted by the International Organization for Migration (IOM) from March 7 to 14, 2024, via phone interviews with South Sudanese returnees from Sudan. Respondents had received transportation assistance from IOM, in coordination with the Government of South Sudan, and were returning to one of four key states: Upper Nile, Unity, Central Equatoria, or Northern Bahr el Ghazal. (<https://dtm.iom.int/datasets/sudan-crisis-returnee-reintegration-perceptions-survey-march-7-14-2024>)

The survey captured returnees’ experiences with reintegration, focusing on access to services like housing, food, healthcare, education, and their perceptions of safety and livelihoods. The data provides valuable insights into early resettlement challenges and available support systems.

For this codebook, the data was refined into a specialized dataset that features the experiences of girls and the risks they face during reintegration. It highlights areas such as education access and protection risks aiming to guide child-sensitive programs and policies.

Variables include indicators related to food and water access, school enrollment, and gender-specific vulnerabilities, which offers insights into educational outcomes and the likelihood of continued education for girls.

Sources and Methodology

The data used in this analysis comes from the *Sudan Crisis – Returnee Reintegration & Perceptions Survey*, conducted by the International Organization for Migration (IOM) between March 7–14, 2024. This original dataset has **159 observations and 303 variables**.

The original dataset from the IOM Sudan Crisis – Returnee Reintegration & Perceptions Survey was filtered to focus on variables relevant to returnee demographics, child well-being, education, and safety. It is important to note that the age of the respondent was not kept due to the original survey’s lack of data having one vague age range of 15-59 category for only ten respondents. Therefore, this demographic variable could not be utilized. To avoid confusion, the respondents of the survey are not young girls but rather head of household and family members (which could include young girls over 15 years old) and their perspectives.

Data cleaning steps included:

- Renaming variables to tidy, readable formats without spaces
- Dropping additional irrelevant or duplicate columns
- Converting key variables—such as number of children and gender—into appropriate numeric formats for analysis
- Re-coding several variables from survey questions that had multiple answers for clarity

This refined dataset, `df_Sudan`, is now structured to support analysis of child-related vulnerabilities, with a focus on girls and education outcomes during reintegration. The cleaned dataset has **159 observations** and **47 variables**.

Citation

- International Organization for Migration (IOM). (2024, March 22). *DTM Sudan Crisis – Returnee Reintegration & Perceptions Survey (March 7–14, 2024)*. IOM South Sudan. (<https://dtm.iom.int/datasets/sudan-crisis-returnee-reintegration-perceptions-survey-march-7-14-2024>)

Variable Description

State Code

- **Variable Name:** `admin1_pcode`
- **Type:** Character
- **Description:** Administrative level 1 code (state) of current location in South Sudan.

County Code

- **Variable Name:** `admin2_pcode`
- **Type:** Character
- **Description:** Administrative level 2 code (county) of current location in South Sudan.

IOM Month of Registration

- **Variable Name:** `month_registered_iom_ota`
- **Type:** Character
- **Description:** Month when the returnee was registered by IOM for onward transportation assistance.

Respondent gender

- **Variable Name:** `gender_respondent`
- **Type:** Numeric (0 = Male, 1 = Female)
- **Description:** Gender of the survey respondent.

Value	Label	Freq
0	Male	98
1	Female	61

Education Level

- **Variable Name:** `education`
- **Type:** Numeric
- **Description:** Highest level of education attained by the respondent.

Label	Freq
No formal schooling	47
Started but did not complete primary school	48
Finished primary school, but did not start secondary school	13
Started but did not complete secondary school	16
Finished secondary school	22
University	13

State of origin (South Sudan)

- **Variable Name:** `origin_state_ss`
- **Type:** Character
- **Description:** Returnee's state of origin in South Sudan.

County of origin (South Sudan)

- **Variable Name:** `origin_county_ss`
- **Type:** Character
- **Description:** Returnee's county of origin in South Sudan.

Payam of origin (South Sudan)

- **Variable Name:** `origin_payam_ss`
- **Type:** Character
- **Description:** Returnee's payam (sub-county) of origin in South Sudan.

State of residence before return (Sudan)

- **Variable Name:** `residence_state_sudan`
- **Type:** Character
- **Description:** State where the respondent lived in Sudan before returning to South Sudan.

County of residence before return (Sudan)

- **Variable Name:** `residence_county_sudan`
- **Type:** Character
- **Description:** County where the respondent lived in Sudan before returning.

Current state of residence (South Sudan)

- **Variable Name:** `current_state_ss`
- **Type:** Character
- **Description:** Respondent's current state of residence in South Sudan.

Current county of residence (South Sudan)

- **Variable Name:** `current_county_ss`
- **Type:** Character
- **Description:** Respondent's current county of residence in South Sudan.

Current payam of residence (South Sudan)

- **Variable Name:** current_payam_ss
- **Type:** Character
- **Description:** Respondent's current payam (sub-county) in South Sudan.

Distance to Nearest Market

- **Variable Name:** distance_market
- **Type:** Factor
- **Description:** Distance to the nearest market from current location.

Value	Label	Freq
1	0-15 minutes	65
2	15-30 minutes	63
3	30-45 minutes	20
4	45 min-1 hour	7
5	1-2 hours	3
6	more than 2 hours	1

Meals Per Day

- **Variable Name:** meals_daily
- **Type:** Factor
- **Description:** Number of meals eaten per day reported by the household.

Value	Label	Freq
0	Less than one meal per day	33
1	One	54
2	Two	65
3	Three or more	2
NA	NA	5

Access to Water

- **Variable Name:** water_access_num
- **Type:** Numeric
- **Description:** Availability or ease of access to water in the current location.

Value	Label	Freq
1	Yes	139
2	No	4
3	Sometimes/inconsistently	16

Number of Children in Household

- **Variable Name:** num_children
- **Type:** Numeric
- **Description:** Total number of children in the respondent's household.

num_children	Freq
0	128
1	13
2	4
3	7
4	4
5	1
6	1
8	1

Number of Boys Traveling

- **Variable Name:** boys_under_18
- **Type:** Numeric
- **Description:** Total number of boys under 18 that family members traveled with.

boys_under_18	Freq
0	40
1	44
2	40
3	20
4	8
5	4
6	2
7	1

Number of Girls Traveling

- **Variable Name:** girls_under_18
- **Type:** Numeric
- **Description:** Total number of girls under 18 that family members traveled with.

girls_under_18	Freq
0	35
1	44
2	43
3	23
4	9
5	3
6	1
8	1

Perception of Safety Upon Return

- **Variable Name:** `safety_ss`
- **Type:** Factor
- **Description:** Respondent's perception of safety in their current location in South Sudan.

Value	Label	Freq
1	Very unsafe	1
2	Unsafe	8
3	Neither safe nor unsafe	1
4	Somewhat safe	48
5	Very safe	98
NA	NA	3

Reasons Location Feels Unsafe

- **Variable Name:** `reasons_unsafe`
- **Type:** Character
- **Description:** Specific reasons given by the respondent for feeling unsafe.

Label	Freq
NA	150
Due to conflict among armed groups	4
Due to localized gang violence	2
Due to conflict, gang violence, and gender-based violence	1
Due to discrimination and cattle-related violence	1
Due to gender-based violence and discrimination	1

School Attendance in Sudan

- **Variable Name:** `child_school_sudan`
- **Type:** Numeric
- **Description:** Whether the children in the household attended school while in Sudan.

Value	Label	Freq
1	Yes	127
2	No	22
3	Some of them, but not all	8
4	I don't have school-aged children	2

Protection Risks for Children

- **Variable Name:** `risks_children`
- **Type:** Character
- **Description:** Different types of risks that children may face, such as abduction/trafficking, forced recruitment, abandonment, exploitation, or GBV. The variable is a character string with responses from multiple risk categories, and some respondents provided combinations of risks.

Notes: -Abduction//Trafficking -Forced recruitment, -Abandonment, -Exploitation, -GBV, -Other, -None

Risk of Abduction/Trafficking

- **Variable Name:** Abduction_Trafficking
- **Type:** Numeric
- **Description:** The risk of abduction or trafficking faced by children, as reported by the respondent.

Value	Label	Freq
0	Risk Not Identified	157
1	Risk Identified	2

Risk of Abandonment

- **Variable Name:** Abandonment
- **Type:** Numeric
- **Description:** The risk of children being abandoned by their caregivers or family members, as reported by the respondent.

Value	Label	Freq
0	Risk Not Identified	146
1	Risk Identified	13

Risk of Forced Recruitment

- **Variable Name:** Forced_Recruitment
- **Type:** Numeric
- **Description:** The risk of forced recruitment of children into armed groups or forces, as reported by the respondent.

Value	Label	Freq
0	Risk Not Identified	154
1	Risk Identified	5

Risk of Exploitation

- **Variable Name:** Exploitation
- **Type:** Numeric
- **Description:** The risk of exploitation of children, as reported by the respondent.

Value	Label	Freq
0	Risk Not Identified	139
1	Risk Identified	20

Risk of Gender-Based Violence (GBV)

- **Variable Name:** GBV
- **Type:** Numeric
- **Description:** The risk of gender-based violence (GBV), as reported by the respondent.

Value	Label	Freq
0	Risk Not Identified	137
1	Risk Identified	22

Other Risks

- **Variable Name:** Other
- **Type:** Numeric
- **Description:** The risks not covered by the other categories, as identified by the respondent.

Value	Label	Freq
0	Risk Not Identified	158
1	Risk Identified	1

Risk of None

- **Variable Name:** None
- **Type:** Numeric
- **Description:** Whether the respondent reported that there were no risks faced by children in their area.

Value	Label	Freq
0	Risks Identified	65
1	No Risks Identified	94

Risk Response of “I choose not to answer”

- **Variable Name:** No_Answer
- **Type:** Numeric
- **Description:** Whether the respondent chose not to provide an answer regarding the risks children face.

Value	Label	Freq
0	Response Given	139
1	Response Not Given	20

Current School Attendance

- **Variable Name:** children_current_school
- **Type:** Numeric
- **Description:** Whether the children in the household are currently attending school.

Value	Label	Freq
1	Yes	58
2	No	75
3	Some of them, but not all	24
4	I don't have school-aged children	2

Reasons for Not All Children Attending School

- **Variable Name:** reason_not_all_in_school
- **Type:** Character
- **Description:** Respondents' open-ended answers explaining why not all school-aged children in their household are attending school. Responses may include one or more reasons, such as inability to afford school fees, distance to school, child labor needs, or other personal or logistical barriers. Multiple responses may be recorded in a single entry.

Reason: Can't Afford Fees / Scholastic Materials

- **Variable Name:** Reason_Fees
- **Type:** Numeric
- **Description:** Indicates whether the respondent mentioned inability to afford fees or scholastic materials as a reason for not all children attending school.

Value	Label	Freq
0	Not Mentioned	21
1	Reason Mentioned	78
NA	NA	60

Reason: No School Nearby / Distance Too Far

- **Variable Name:** Reason_Distance
- **Type:** Numeric
- **Description:** Indicates whether the respondent mentioned distance or lack of nearby schools as a reason for not all children attending school.

Value	Label	Freq
0	Not Mentioned	93
1	Reason Mentioned	6
NA	NA	60

Reason: Children Need to Work

- **Variable Name:** Reason_Work
- **Type:** Numeric
- **Description:** Indicates whether the respondent mentioned the need for children to work as a reason for non-attendance.

Value	Label	Freq
0	Not Mentioned	98
1	Reason Mentioned	1
NA	NA	60

Reason: Other

- **Variable Name:** Reason_Other
- **Type:** Numeric
- **Description:** Indicates whether the respondent selected “Other” as a reason for non-attendance, which may include personal, cultural, or unspecified reasons.

Value	Label	Freq
0	Not Mentioned	80
1	Reason Mentioned	19
NA	NA	60

Reason: Chose Not to Answer

- **Variable Name:** Reason_NoAnswer
- **Type:** Numeric
- **Description:** Indicates whether the respondent selected “I choose not to answer” when asked why not all children were attending school.

Value	Label	Freq
0	Not Mentioned	94
1	Reason Mentioned	5
NA	NA	60

Protection Risks for Women and Girls

- **Variable Name:** risks_women_girls
- **Type:** Character
- **Description:** Different types of risks faced by women and girls during reintegration such as abduction/trafficking, forced recruitment, abandonment, exploitation, or GBV. The variable is a character string with responses from multiple risk categories, and some respondents provided combinations of risks.

Risk of Abduction/Trafficking for Women and Girls

- **Variable Name:** Abduction_Trafficking_women
- **Type:** Numeric
- **Description:** Whether the respondent identified that women and girls face a risk of abduction or trafficking.

Value	Label	Freq
0	Risk Not Identified	158
1	Risk Identified	1

Risk of Abandonment for Women and Girls

- **Variable Name:** Abandonment_women
- **Type:** Numeric
- **Description:** Whether the respondent identified that women and girls face a risk of abandonment.

Value	Label	Freq
0	Risk Not Identified	156
1	Risk Identified	3

Risk of Forced Recruitment for Women and Girls

- **Variable Name:** Forced_Recruitment_women
- **Type:** Numeric
- **Description:** Whether the respondent identified that women and girls face a risk of recruitment into armed groups or forces.

Value	Label	Freq
0	Risk Not Identified	157
1	Risk Identified	2

Risk of Exploitation for Women and Girls

- **Variable Name:** Exploitation_women
- **Type:** Numeric
- **Description:** Whether the respondent identified that women and girls face a risk of exploitation.

Value	Label	Freq
0	Risk Not Identified	141
1	Risk Identified	18

Risk of Gender-Based Violence (GBV) for Women and Girls

- **Variable Name:** GBV_women
- **Type:** Numeric
- **Description:** Whether the respondent identified that women and girls face a risk of gender-based violence (GBV).

Value	Label	Freq
0	Risk Not Identified	104
1	Risk Identified	55

Other Risks for Women and Girls

- **Variable Name:** Other_women
- **Type:** Numeric
- **Description:** Whether the respondent identified any other risks not covered by the other categories for women and girls.

Value	Label	Freq
0	Risk Not Identified	158
1	Risk Identified	1

No Risks for Women and Girls

- **Variable Name:** None_women
- **Type:** Numeric
- **Description:** Whether the respondent identified that there were no risks faced by women and girls in their area.

Value	Label	Freq
0	Risks Reported	86
1	No Risks Reported	73

No Answer for Women and Girls

- **Variable Name:** No_Answer_women
- **Type:** Numeric
- **Description:** Whether the respondent chose not to answer regarding the risks faced by women and girls.

Value	Label	Freq
0	Answer Provided	140
1	No Answer Provided	19