

---

Subject: gender matching interviewer / participant in Zimbabwe & Burundi

Posted by [jwilliamrozelle](#) on Sat, 02 Oct 2021 10:35:43 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

I've been digging into the Fieldworker datasets, and merging datasets based on v028 and fw101.

Based on this join, it looks like Burundi (BU7) has 276 men's questionnaires administered by females, and zimbabwe (ZW7) has 186 men's questionnaires administered by women. Then, Zimbabwe looks like it also has 71 of the women's questionnaires administered by a male.

Is this correct, or am I doing something wrong here?

I couldn't find information about it in the reports, but maybe I'm missing it.

Thanks!

---

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi

Posted by [Bridgette-DHS](#) on Mon, 04 Oct 2021 20:36:20 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The problem is that when you merged the FW and IR files, (and the FW and MR files) you should have dropped the interviewers who did not match with a woman in the IR file (or a man in the MR file). The following Stata code will work for the FW and IR merge with BU7:

```
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\BUFW71FL.DTA", clear
keep fw101 fw105
rename fw101 interviewer_id
rename fw105 interviewer_sex
sort interviewer_id
save tempfw.dta, replace
```

```
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\BUIR71FL.DTA", clear
keep v028
rename v028 interviewer_id
sort interviewer_id
merge interviewer_id using tempfw.dta
```

```
tab _merge interviewer_sex
keep if _merge==3
tab interviewer_sex
```

I use an older version of the merge command, but whatever version you use, you need to drop the cases for which `_merge` is not 3.

Good to see that you are using the FW files!

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi

Posted by [jwilliamrozelle](#) on Mon, 04 Oct 2021 23:19:18 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

First of all, huge thanks to Tom and Bridgette for the really quick response. Tom, I've read through your interviewer effects report and nerded out about it - love the work.

Normally I use R, but I pulled up STATA and followed your code - and still ended up with the same number of unmatched gender in the men's recode Burundi dataset (BU7), and in the men and women's recode in Zimbabwe (ZW7). See here in ZW7, 71 observations with male interviewers. See the tab results below (prior to dropping everything but \_merge3)

```
. tab _merge interviewer_sex
```

| _merge | male | female | Total  |
|--------|------|--------|--------|
| 2      | 49   | 1      | 50     |
| 3      | 71   | 9,884  | 9,955  |
| Total  | 120  | 9,885  | 10,005 |

Additionally - the IR dataset starts with 9955 observations, which is the same number of observations I have in my final R dataframe (and also the merge==3 dataset)

To narrow this down further, it looks like there are observations from interviewers 201, 207, 605, 606, 820, 904 and 926 who are male.

```
tab interviewer_id interviewer_sex if interviewer_sex == 1
```

| interviewer_id | male | Total |
|----------------|------|-------|
| 121            | 2    | 2     |
| 200            | 2    | 2     |
| 201            | 33   | 33    |
| 207            | 24   | 24    |
| 400            | 1    | 1     |

|                   |    |    |
|-------------------|----|----|
| 605               | 1  | 1  |
| 606               | 2  | 2  |
| 820               | 2  | 2  |
| 904               | 1  | 1  |
| 926               | 3  | 3  |
| -----+-----+----- |    |    |
| Total             | 71 | 71 |

Similarly, for Burundi men's recode (prior to dropping everything but \_merge==3):

```
tab _merge interviewer_sex
```

| _merge            | fieldworker sex |        | Total |
|-------------------|-----------------|--------|-------|
|                   | male            | female |       |
| 2                 | 34              | 92     | 126   |
| 3                 | 6,738           | 276    | 7,014 |
| -----+-----+----- |                 |        |       |
| Total             | 6,772           | 368    | 7,140 |

But in Burundi, it seems that it was all one enumerator (id 2104)

```
tab interviewer_id interviewer_sex if interviewer_sex == 2
```

| interviewer_id    | fieldworker sex |  | Total |
|-------------------|-----------------|--|-------|
|                   | female          |  |       |
| 2104              | 276             |  | 276   |
| -----+-----+----- |                 |  |       |
| Total             | 276             |  | 276   |

Also, as long as I'm connected with you - had a few quick questions about some missingness in interviewer info. I'm picking up some missing - and Tom, looks like you had the same issue in the methodological report based on table 2.2.

Curious if there's a broad sense about why that's missing in some countries - but also understand

that there's a degree of removal from the actual survey implementation.

Thanks again!

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi  
Posted by [Bridgette-DHS](#) on Wed, 06 Oct 2021 11:35:01 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

Thank you for the positive feedback on the interviewer effects work. Much appreciated!

You are right that the Zimbabwe survey indicates that men were interviewing women and women were interviewing men, and this is not explained in the same way as in the Burundi survey. I ran the following Stata lines to get lists of which interviewer codes were involved (I know you use R but this is easier for me):

\* Zimbabwe

```
cd e:\DHS\DHS_data\scratch
```

```
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\ZFW71FL.DTA", clear  
keep fw101 fw105  
rename fw101 interviewer_id  
rename fw105 interviewer_sex  
sort interviewer_id  
save tempfw.dta, replace
```

```
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\ZWIR72FL.DTA", clear  
keep v001 v002 v003 v012 v024 v025 v028 v030  
rename v028 interviewer_id  
sort interviewer_id  
merge interviewer_id using tempfw.dta  
tab _merge interviewer_sex  
keep if _merge==3  
tab interviewer_sex  
tab interviewer_id if interviewer_sex==1
```

```
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\ZWMR72FL.DTA", clear  
keep mv001 mv002 mv003 mv012 mv024 mv025 mv028 mv030  
rename mv028 interviewer_id  
sort interviewer_id  
merge interviewer_id using tempfw.dta  
tab _merge interviewer_sex  
keep if _merge==3  
tab interviewer_id if interviewer_sex==2
```

I suspect that all of these irregular cases can be explained by a combination of coding errors in

the FW data or misuse of interviewer ID codes. I asked Trevor Croft and he gave the following detailed interpretation. Nevertheless, this issue may complicate your analysis. You may want to simply drop these cases from your analysis. At least you can be sure that the inconsistencies were not edited out, giving a false impression that the data were perfect!

From Trevor Croft:

Quote:Here is what I could figure out:

Interviewer 803 only did men's interviews and is almost certainly a man. This looks like a coding error for the interviewer's sex.

Interviewer 920 is a supervisor code, and it appears that the person using the code at the beginning of the survey was female, but left the interviewing teams. The replacement supervisor was male and was originally interviewer 903 in a different team. 903 worked as a male interview in July and the beginning of August, but then was supervisor (using code 920) but was also conducted some men's interviews. This is not what should have happened. It can be corrected by recoding interviewer code 920 in the men's data as code 903.

Interviewers 201 and 207 are not clear. I think that there was probably some confusion in the team in terms of interviewer numbers and for a couple of clusters each a female interviewer used a male interviewer's number.

It is definitely possible that v028 and mv028 may be incorrect. Interviewers starting data collection in the cluster first provide the team number, then from within a list for the team they select their interviewer number. It is almost impossible for them to get the team number wrong as many things would fail, but it is definitely possible for them to confuse interviewer numbers, and we see a smattering of other cases that appear to be that type of problem. For 201 and 207 it appears that these were used incorrectly for two clusters each. It is possible that a couple of temporary interviewers were given the codes 201 and 207. I'm guessing for these two.

I might just mention that historically a few surveys have permitted men to interview women and women to interview men. For example the first Ghana DHS in 1988 permitted that as it was standard practice for their surveys at the time.

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi  
Posted by [jwilliamrozelle](#) on Thu, 07 Oct 2021 03:04:59 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Bridgette, Tom, and Trevor

Thanks again! Deeply appreciate that you folks go above and beyond to make this data accessible and usable for people like me. You all are great!

I actually didn't have issues with Burundi IR, just the MR. I think there's still an issue with

interviewer 2104 as a female interviewing males. I could probably check to determine whether that ID shows up in the IR.

In terms of analysis - the gender itself shouldn't matter much - as it seems as though I can safely assume that I've been merging the same way you would and that in practice, all datasets with the FW data available gender matched interviewers / participants. Other variables, like age, education, etc. do matter a bit. From a combination of Tom's and Trevor's response... sounds like I should:

- Recode enumerator 920 in ZW7 MR from 920 to 903
- Keep interviewer 803 (assuming that the FW data, except for interviewer sex is probably correct)
- Drop the rest (primarily only meaningful for 201 and 207) as it doesn't sound like it's very safe to assume that the FW data is for the people who did the interviews.

No worries at all about STATA, I appreciate the opportunity to dust off the STATA license on my computer. I really should be better about maintaining that skillset. I did notice that DHS program has started to publish code to reproduce indicators on github, including in R. Very cool! I'm not sure if you take pull requests from the public - but if you do, would be happy to submit some pull requests following the conventions you've got there for some of the sections that haven't been published yet.

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi  
Posted by [Bridgette-DHS](#) on Fri, 08 Oct 2021 12:32:52 GMT  
[View Forum Message](#) <> [Reply to Message](#)

Following is another response from DHS Research & Data Analysis Director, Tom Pullum:

Glad to help. Regarding possible fixes--you can make those changes, to engineer better consistency among the responses, but we can't be certain that they are corrections....

Now that we have timestamps, it's possible to track interviewers in great detail as they move through a cluster and from one cluster to the next. Fascinating--for some of us, at least!

Yes, our goal is to have programs on GitHub in Stata, SPSS, and R for all the chapters, including extra chapters for modules that are only occasionally used. Certainly, any suggestions or pull requests will be welcome!

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi  
Posted by [jwilliamrozelle](#) on Thu, 23 Dec 2021 03:44:37 GMT  
[View Forum Message](#) <> [Reply to Message](#)

Hi again, I'm back with what (I hope) are the last question on this.

I wanted to integrate the 2015 AIS from Mozambique (MZ6) into the analysis. I'm seeing from the documentation that to match for interviewer ID, I should add 1000 to the v805/mv805 (actually, I've been using (m)v028 - but it looks like it's the same). That's leaving me with 848 (10.9%) of observations without a matched interviewer in the individual recode, and 8.3% unmatched in the Men's recode. Is there any additional guidance on how I might merge these datasets?

Thanks again.

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi  
Posted by [Bridgette-DHS](#) on Fri, 24 Dec 2021 17:38:22 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is another response from DHS Research & Data Analysis Director, Tom Pullum:

It appears that the fieldworker ID code (fw101) in MZFW71 is inconsistent with v028 in MZIR71 and mv028 in MZMR71. It is not as simple as a difference of 1000. Will get back to you after resolving this inconsistency.

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi  
Posted by [Bridgette-DHS](#) on Tue, 28 Dec 2021 23:02:06 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is another response from DHS Research & Data Analysis Director, Tom Pullum:

I have found (thanks to to Ruben Hume of the Data Processing staff) that MZ71 included some survey-specific codes for interviewers, etc. In the IR file, the health technician was identified with sbtecnum4. (In general, look for variables that contain "num4" in their name.) In the MR file, the variable is smbtecnum4. I see that these are 4-digit codes that begin with 4, specifically 4001 through 4050, and they match with fieldworkers 4001 through 4050 in the FW file. The following Stata code will do the complete merge. I will also paste the 2x2 table for sex of respondent vs sex of interviewer for the merged cases. It shows that there is no relationship between these two variables. Note that by redefining "int\_id" to be one of the other interviewer codes you can link to other 4-digit codes in the FW file.

```
cd e:\DHS\DHS_data\scratch
```

```
* Prepare the IR file  
use "...MZIR71FL.DTA", clear  
gen cluster=v001  
gen hh=v002
```

```

gen line=v003
*gen int_id=v028
gen int_id=sbtecnum4
keep cluster hh line int_id
gen sex=2
save IR.dta, replace

* Prepare the MR file
use "...MZMR71FL.DTA", clear
gen cluster=mv001
gen hh=mv002
gen line=mv003
*gen int_id=mv028
gen int_id=smbtecnum4
keep cluster hh line int_id
gen sex=1
save MR.dta, replace

* Prepare the AR file
use "...MZAR72FL.DTA", clear
gen cluster=hivclust
gen hh=hivnumb
gen line=hivline
gen result=hiv03
sort cluster hh line
save AR.dta, replace

* Prepare the FW file
use "...MZFW71FL.DTA", clear
gen int_id=fw101
gen int_sex=fw105
keep int_id int_sex
sort int_id
save FW.dta, replace

* Append IR and MR
use IR.dta, clear
append using MR.dta
sort cluster hh line

* Merge with AR
merge cluster hh line using AR.dta
rename _merge AR_merge
sort int_id

* Merge with FW
merge int_id using FW.dta
rename _merge FW_merge

```



\* Compare sex of interview and sex of respondent  
tab sex int\_sex,m  
tab sex int\_sex if FW\_merge==3

---

### File Attachments

1) [merge.png](#), downloaded 130 times

---

---

Subject: Re: gender matching interviewer / participant in Zimbabwe & Burundi  
Posted by [jwilliamrozelle](#) on Tue, 28 Dec 2021 23:48:05 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Amazing! As always - I'm grateful for your responsiveness and support.

---