

Data Outline—Chase Africa

Contents

1	Files and sheets	1
1.1	Tables	1
1.1.1	Row (case)–	1
1.1.2	Column (variables):	1
1.2	Data janitor work	3
1.2.1	Dandelion	3

1 Files and sheets

Six files of annual and all year data summaries for six different projects?

- 2012-18 CHAT data summary CP.xlsx – 7 annual sheets + summary sheet
- 2014-18 Dandelion data summary & CYP CP.xlsx– 5 annual sheets + summary sheet
- 2014-18 MKT data summary & CYP CP.xlsx – 5 annual sheets + summary sheet
- Big Life data summary & CYP CP.xlsx – 2 annual sheets + summary sheet
- CHV data summary & CYP CP.xlsx - 3 annual sheets + summary sheet
- RICE WN Data summary & CYP CP.xlsx– 1 annual sheet + summary sheet

The MKT data sheets additionally have some vehicle contribution payment schedules attached

The CHV data sheets additionally have a pre-chase data sheet?

Overall structure seems similar: over the years the tables become more detailed, include more columns, the final ones are most complete.

Some calculations are made using formulas in the sheets, not all though.

1.1 Tables

1.1.1 Row (case)–

Rows do not correspond to unique cases, which would ideally be individual clinic and date combination, so the data is not *tidy*, but instead switch between cases and (sub-)totals, with some money related cells having an ad hoc structure as well. It all seems readable, but is a dangerous precedent that could easily lead to errors.

1.1.2 Column (variables):

1.1.2.1 Location and funding of free clinic

- *date of day-clinic*
 - *venue*
 - (also additional information about cost sharing?)
- *amount / date funding received*
 - there are amounts in Ksh and £, additionally also apportionment of money sometimes?
 - if there are two or three rows per ‘case’, not clear if this is to do with the individual grants?
- Does this mean there are funding periods for which the grants apply?

1.1.2.2 Mobilization, Education

This category of variables seems to only exist in the CHAT files.

- *Mobilization, education* (until 2017)
- sometimes split (Male, Female, Total)
- *Behaviour change (Sensitization)* 2018, but classed as Mob-ed in summary tables?
 - split (Male, Female, Total)
- *Youth out of school sensitised on RH/FP Mobilization, ecological sensitization* (until 2015)
 - split (Male, Female, Total)

1.1.2.3 Family planning

- *LONG TERM METHODS*
 - divided into IUCD, 5yr, 5 yr rep, 3 yr and tubal ligation
 - sometimes data not known and “5/3yr split not provided so 50:50 assumed”
- *SHORT TERM METHODS*
 - divided into Depo and pills (6, 3, 1months)
 - sometimes just says “Pills”, not clear how long?
 - Also “pills No. new” and “pills no.rev”, but these seem to be counted differently? in the CYP calculations?
 - *AGE RANGE OF CLIENTS (FP)* (from 2016)
 - in 2016 and 2017 0-15/ 16-30/ over 30
 - in 2018 under 18 and over 18
- *TOTAL FP* - total family planning
 - derived = sum of long and short term methods
- *CYP TOTAL* - couple years of protection
 - derived from long and short term methods and a set of CYPs e.g. 3.8 for 5 year implant = these are all contained in the formulas!
 - IUCD 4.6
 - 5 year implant/rep = 3.8
 - 3 year implant/rep = 2.5
 - Depo 1st/ Depo rep = 0.25
 - Pills (3 mths) = 0.2
 - Pills 6 mnth = $1/2.2 = 0.4545$
 - Pills 1 mnth = $1/15 = 0.666$
 - nothing on the other two pillse oclumns 'no. new and no. rev“?
- *cost per person* and *Cost per CYP* derived from dividing amount by FP total and CYP total
- *CONDOMS* - in guess units? seems to be in all the tables
 - not used in CPY calculations
- *Implant removal*
 - This only turns up in 2018 I think. No info on continuation or not?
- *persons with disability FP*
 - only 2018 I think

CYP measures the volume of programme activity from service statistics or logistics management information systems. It is currently the most widely used indicator of output in international family planning programmes. It's easy to calculate from routinely collected data.

Issues:

- Does not give any indication of the number of individuals affected/represented by the number.
- conversion factors are debatable?
- reflects distribution and not actual use or impact?
- you “credit” the entire amount to the year in which the method was delivered. e.g. the 9 year CYP for a sterilization will count in that single year, even though the protection will be realised over the next 9

years. *annualization* is an option though, we can split it over the next nine year period! This should happen with all long term contraception

- what about removal? it's part of access to FP services, yet reduces CYP. But doesn't seem to get included in calculations.

1.1.2.4 Integrated health services (NA 2011/12)

- *Primary health care* or sometimes *Basic health care* (NA 2011/12)
- *TB* - only in CHAT, 2014 and 2015 I think
 - *TB - contacts traced*
 - *TB household education*
 - **TB - defaulters traced**
- *HIV/AIDS*
 - *TEST*
 - *+ve*
- *Deworming children* (NA 2011/12, 2013, 2014, 2015)
 - *Immunization* (NA 2011/12, 2013, 2014, 2015)
- *malaria test* test/+ve (only 2018)
 - *TEST*
 - *+ve*
- *cancer screening* test/+ve(only 2018)
 - *TEST*
 - *+ve*
- *hep B test* test/+ve (only 2018)
 - *TEST*
 - *+ve*
- *persons with disability health services* (only 2018)
- *Health services only* - derived from primary health care, TB contacts traced and HIV tests, and deworming and immunization when available - sum of people treated?
- *Health & Family planning* - derived sum total family planning and health services.

1.2 Data janitor work

1.2.1 Dandelion

Looks overall to be a consistently processed worksheet:

- 2014-2018 i.e. five worksheets with same number of columns
- subtotals and totals need to be removed
- funding variable needs to be manually split into three variables
- In 2015 sheet, the last funding period date also says "PopOff" in the same cell.
- In the 2016 sheet, the last funding period doesn't have a date, but has a euro value instead. . .
- I remove the euro value and have a missing date in 2016. Actually no, I'll remove this whole funding period "Amplify change". Actually, no, I've added it back and added a column on "fund_category" distinguishing between standard and others.
- In 2017 there an Amboseli vs standard clinic distinction. Annoyingly one of them only has 2 rows, but the funding data is in three..
- ERROR: In 2018 the Pills column is split into 6, 3, or one month. Looking at the CYP factors in the previous sheets it looks like the previous columns were for single month? BUT that means the "All years" total sheet has it wrong, since it adds the 2018 data into the same column as the previous years, even though in 2018 it's all 3 month. So either it should be in a different column, or the previous four sheets give the pill the wrong factor!

- I will assume that the pills for the first four sheets are monthly.
- QUESTION: in 2019 there is a new column “Basic Healthcare”, which is probably the same as “Primary Healthcare” in the previous sheets? I will assume that it is.
- Also in 2018 Now cancer is split into test and positive. I am assuming that the previous years was testing, not another variable?
- Some cells in the excel tables are empty instead of having zeros. Are these different? I will treat them as zeros here, not NAs, but this needs to be confirmed.
- Formula copy paste error: 2016 the CYPs from August onwards do not take into account IUCDs.
- More error: In all of 2017 the IUCD are counted as 3.3 instead of 4.6 as they are everywhere else.
- Inconsistency - order of last two columns in 2018 is switched compared to others. Not error though.
- Error: in 2015 the total health care formula included the HIV positive people i.e. double counted them.
- Because there is a missing funding date in the data, I’ll just make one up !!! as 1.9.2016 for the amplify change one.