NRVA 2003 – Metadata

1 NRVA 2003 files

All the NRVA files have been grouped together into 4 zip files and posted on the website. These are:

Zip File Name	Files contents
NRVA 2003 Documentation. zip	All documentation needed, including this metadata document; • NRVA 2003 Metadata.doc • All the questionnaires and guidelines. • NRVA 2003 A Stakeholder Generated Methodology-AREU.pdf, a report on the development of the methodology. • NRVA 2003 Consumption data cleaning to medians.xls - Documentation on cleaning of food-consumption data from household questionnaire. • NRVA 2003 Data request form.doc • NRVA 2003 Food kcalorie values.xls - Food calorie values used for calculations.
NRVA 2003 District data.zip	• NRVA 2003 Ag-Ecological Zone District data.xls - Data on how agro-ecological zones within districts collected at district government level. • NRVA 2003 Kuchi District data.xls- Kuchi information from district government level. • NRVA 2003 District price data.xls - District level prices.
NRVA 2003 Shura & Wealth Group data.zip	• NRVA 2003 Male Female Shura.xls - Data from village councils (shuras) both male and female. • NRVA 2003 Male Wealthgroup.xls - Data from male wealth groups. • NRVA 2003 Female Wealthgroup.xls - Data from female shura wealth group.
NRVA 2003 Household data-Excel-DBF.zip	• NRVA 2003 Household data.xls - Household level data in Excel format. • NRVA 2003 HH Member data.dbf - Household member level data in DBF format.

2 Strata

The methodology, questionnaires and guidelines for the NRVA 2003 are available on the website www.af/NRVA, but also posted on the download web site in the file ***NRVA 2003 Household Questionnaires & Guidlines.zip***.

Data were collected from all provinces, and districts within provinces, except for 11 districts, where was not collected because of security restrictions. Data were collected at the following units of observation with the following linking variables indicated.

Unit observation	# Observations	Relational linking variable	Data release format(s)
Districts	368		
Villages or communities with villages	1,853	NRVA District Code	Excel
Wealth groups within villages/communities (3 per village)	5,559	Shura Code	Excel
Households within wealth group	11,765	Wealth Group Code	Access, Excel
Households within wealth group (Assets & Shocks data)	11,706	Wealth Group Code	Access, Excel
Households within wealth group (Land tenure, Crops & Livestock data)	10,831	Wealth Group Code	Access, Excel
Households within wealth group (Dietary diversity data)	11,227	Wealth Group Code	Access, Excel
Household members (Register, Migration, Education, Health & Employment)	85,577	Household Code or Qaire_ID	Access, DBF

For the smaller data sets (district, shura and wealth groups), it has been possible to output from the Access databases to Excel. A single column is outputted for each variable with the question included in the first cell of the column, with the second often being the abbreviated variable name, more appropriate for import into a statistical package. A comment on the first cell in the column indicates, where appropriate, the code meanings. At the household member level, Excel's worksheet maximum capacity (worksheet size maxim on = 65,536 rows * 256 columns) has been exceeded and therefore the household member data has been exported to a DBF files as well as available in the original Access database.

2.1 Coding system

The coding system used progressively includes more information from the highest level of province right the way through to household member following the scheme laid out in the table below. Strata from shura and below also have included Stratum ID, which are just the additional digits for each code at that strata. This is useful for analysis, for example, to group by wealth group type. For this, Wealth Group Code is not useful as it uniquely at identifiers any single wealth group, but Wealth Group ID is a useful grouping factor.

Stratum code	Coding system
Province Code (2 digits)	1-32 (AIMS 32 District 1982 province model)
AIMS district Code (+2 digits)	Coding of districts within provinces according to the AIMS 1982302 province 329 district model.
NRVA district Code (+2 digits)	Coding of districts within provinces according to the AIMS 1982302 province 329 district model, where they are still applicable. Where they are not, and NRVA code was created along the same scheme, but using a higher number four district within provinces and previously used. For example, if the aim is district code was 1003, but this district had divided into two, the new district might take the code 1023, if there were more than 13 districts in the province, precluding a possibility to use code 1013.
Shura Code / ID (2 digits)	Often and village level, but sometimes it mosque within large village. Extra code runs from one to where *n* is the number of community survey within the district. Maximum value =11.
Wealth group Code / ID (+1 digits)	Additional one digit code refers to either one of the three wealth groups survey it: 1 = medium/better off 2 = poor 3 = very poor. Notice that in the original data, this coding system was reversed in the household data, i.e. the very poor group was coded as one, but this was subsequently been harmonised with the data from the shura and the wealth groups.
Household Code / ID (+1 digit)	Additional one digit code refers to a households surveyed within the wealth group. Maximum value =6.
Individual Code / ID (+2 digits)	Additional to digit code refers to the individual within the household for which information is being collected. Maximum value =20.

2.2 Weighting

The lack of reliable population and village location data ensures that NRVA 2003 could not be designed using a proportional sampling frame. Because of the lack of information on the relative populations of the zones within districts not to mention districts and provinces, the **only truly representative statements that the data can make is about the population sampled**, i.e. 1853 villages, which represent 175,026 households, and 1,273,314 household individuals (assuming average NRVA household size was 7.27) in all but 11 districts. If we assume the rural population is as the latest Central Statistics Office (CSO) population estimates indicates, 16.06m, then this population of individuals represents 12.6% of the Afghan rural population.

Offsetting this lack of statistical representative sampling frame is the large sample size for this type of survey. While we are unclear how to weight between villages, the large sample size gives us a reasonable sense of confidence that the data is probably quite robust. It is hoped that the current CSO pre-census listing accompanied by geo-references will eventually enable a retro weighting to be applied to this sample.

Therefore the weighting system can only accommodate for the different sizes of wealth groups and number of households and individuals interviewed within each wealth group. The weighting system multiplies an individual observation (x) by corresponding weight (w); see comparisons below.

Statistic (x=data; w=weight)	Un-weighted	Weighted
Mean	$\sum x / n$	$\sum wx / \sum w$
Variance	$\Sigma\left(x-\left(\Sigma x/n\right)\right)2 / n-1$	$\Sigma\left(x-\left(\Sigma x/n\right)\right)2 / \sum w-1$

As can be seen from the variance above the weighting introduces an increased ***n***, which is actually pseudo-replication because we do not have **Σ*w*** observations but ***n***.

2.3 Weighting system for Shura level statements (male and female)

The weighting system, for producing means, medians, percentages, etc., where maintaining the correct degrees of freedom is not important. Since one interview was done in each shura (village/community), the weight for each shura interview is simply the number of households in that shura. This system of weights adjusts the sample to be representative of the number of households in the sample.

This weighting system will increase the ***n*** dramatically, creating the pseudo-replication explained above. This is irrelevant when calculating means, medians, percentages, and other descriptive statistics where degrees of freedom are *not* considered, but becomes an issue when using statistical tests where the number of degrees of freedom is important in determining is significance. The way different statistical packages handles such a weighting system may vary.

Note: When drawing conclusions at the shura or community level (e.g.- % of shuras producing handicrafts in the winter) **no** weighting system should be used.

2.4 Weighting system for Wealth groups within village/community statements

Wealth groups within village/community were the next unit of observation, and these should be weighted by their population size. In a wealth group we typified one household to represent the wealth group, and therefore the weight of this wealth group information is equal to the population of that wealth group.

2.5 Weighting system for Household level statements

1) Weighting system, for producing means, medians, percentages, etc., where degrees of freedom are not important.

The weight for each household is calculated using the following formula:

$$\frac{(\#hh\_in\_wealth\_group)}{(\#\_interviews\_in\_wealth\_group)}$$

This system of weights adjusts the sample to be representative of the households the data was sampled from (that is, representative of the villages included in the survey). The sum of these weights is equal to the number of households in all of the villages sampled.

2.6 Weighting system for individual level statements

The following formula should be used with weighting system for households:

$$\frac{(\#hh\_in\_wealth\_group)}{(\#\_interviews\_in\_wealth\_group)} * \#\_members\_in\_hh$$

This is equal to:

$$= HH\_weight * \_\#\_members\_in\_hh$$

2.7 Examples

% of **households** have a kcal/capita lower than basic needs, use the **household weights**, but if you wish to calculate a **headcount rate** of those below the calorific basic requirements, then **individual weights** should be used

% of the **population** with a dietary diversity score greater than 10, use the **individual weights**, because the statements concerns the population of individuals rather than population of households.

3 District data

District data is in the file named :
NRVA 2003 Ag-Ecological Zone District data.xls.

Each row within the table represents an agro-ecological zone within a district. There are a total of 614 such zones in the NRVA 2003. This represents sections A and B from the district questionnaire.

The Kuchi information from the district questionnaire, is contained in the file ***NRVA Kuchi district data.xls.*** Information at a district level for Kuchis was collected only from 80 districts; therefore it is presented in the different spreadsheet to the rest of the district data. The months used in questions are from the Hedri Shamsi Islamic solar Calendar, and the conversion table to the Gregorian calendar is outlined in the table below.

Hedri Shamsi Islamic solar Calendar	1st date	Last date
1. Hamal	21st March	19th April
2. Sawar	20th April	20th May
3. Jawza	21th May	20th June
4. Saratan	21th June	21st July
5. Asad	22nd July	21st August
6. Sunbila	22st August	21st September
7. Mizan	22st September	21st October
8. Aqurab	22st October	20st November
9.Qaus	21st November	20th December

10. Jadi	21th December	19th January
11. Dalwa	20th January	18th February
12. Hut	19th February	20th March

Throughout the data, there is a variable included, **Kuchi_ID**, which is binary variable, with 1 = kuchi community.

District price data is contained within the file ***NRVA District price data.xls.*** This includes all of the price data.

4 Shura of village/community data

Data for male and female shura (councils) is in the file named:
NRVA 2003 Male Female Shura-Excel.zip

5 Wealth group data

The data from the male and female wealth groups is contained in the file is named:
NRVA 2003 Male Wealthgroup-Excel.zip
NRVA 2003 Female Wealthgroup-Excel.zip

Wealth groups were coded in the shura and wealth group questionnaires as:
1. Medium/Better off
2. Poor
3. Very poor

But unfortunately in the field version of the house or questionnaire they were coded in the reverse order:
1. Very poor
2. Poor
3. Medium/Better off.

Therefore some of you may have seen previous versions of the questionnaire, but all data has now been standardised on the wealth group coding used in the original surer and wealth group questionnaires, i.e.:
1. Medium/Better off
2. Poor
3. Very poor.

The copy of the household questionnaire on the down low web site is also being appropriately modified.

6 Household data

The e household data is contained in the files named:
NRVA 2003 Household data-Excel.zip
NRVA Household data Access Database.zip.

The NRVA household level questionnaire was composed of three major sections at the household level: individual member data, assets and shocks, and an agricultural section covering lands, crops and livestock. The questionnaire was designed so these three sections could be managed independently. For several reasons, including the fact that women will not stay able to go to all villages, not all sections of the household questionnaire were answered for all households interviewed. Therefore we have a variable number of responses for the three sections of the household questionnaire (see table above for numbers). Therefore it is necessary to generate three different weighting systems to account for the fact that not all the sections were answered in all of the questionnaires. These weighting systems are included in the appropriate worksheets for the household level data in the file ***NRVA 2003 Household data-Excel.zip***, and are explained below.

Household questionnaire section	Household weighting variable name	Individual weighting variable name
Assets & shocks	HH_WeightAS	Ind_WeightAS
Agriculture (crops, land and livestock)	HH_WeightAG	Ind_WeightAG
Dietary diversity	HH_WeightDD	Ind_WeightDD
Household member data	HH_Weight	Ind_Weight

7 Food-consumption data from household questionnaire

The food consumption data was converted to calories using "as purchased" calorific values quoted in FAO's ***Food Composition Tables for the Near East***. It contains information for foods consumed in the following countries: Afghanistan, Bahrein, Cyprus, Egypt, Iran, Iraq, Jordan, Kuwait, Lebanon, Libya, Oman, Pakistan, Qatar, Saudi Arabia, Somalia, Sudan, Syria and Yemen. It is available online at http://www.fao.org/DOCREP/003/X6879E/Ac. The values from this publication, and those values used for foodstuffs that were not reported in the table but calculated are in the

file ***NRVA 2003 food kcalorie values.xls***. There is a worksheet with the values ready for analysis, and another worksheet with a more detailed explanation of the calculations for some of the non-standard food that values were not available from the tables quoted above.

Egg consumption was reported in the questionnaire as the number of eggs consumed rather than kgs. It is presented in both the Excel and access database as kgs, with the assumption that 1 kg is equal to 20 eggs. Local market prices for the eggs were also collected for individual eggs rather than kgs of eggs, but are presented in the Excel sheet as price per kg to prevent confusion.

"As purchased" calorific values were used because it includes the wastage or refuge % between food purchase and calories utilised. The refuse is the portion of the food as purchased which is commonly not eaten. It may represent an inedible part of the food such as the shell of a nut, or an edible part which may or may not be consumed, such as the peel of an apple. The refuse value of a food item varies depending on the food.

The household level food-consumption form had two questions for each food type:

P2. Has any member of the household consumed [FOOD ITEM] in last 7 days?
P3. How much [FOOD ITEM IN UNITS] was consumed in total by the household during the last 7days?

	Kilo calories	
Age	Male	Female
0-4	1320	1250
5-9	1980	1730
10-14	2370	2040
15-19	2700	2120
20-59*	2460	1990
60+*	2010	1780
Figures given here apply for "light" activity level (♂=1.55 * BMR; ♀=1.56 * BMR)		

In many cases question P2 was answered with a 2 (No), but then the respondents went on to indicate the quantity of the food eaten in the last seven days. The analysis of the consumption data to date has disregarded question P2 data, and taken any reported quantities of foods consumed in question P3. Therefore only weights are presented in the Access and Excel files.

On the food-consumption questionnaire, there was space for 10 other additional foods to be specified and quantified in terms of the consumption for the household over the last seven days. This data has been disregarded as there was so little of it and frequently not accompanied by text identify what food was being referred to.

Consumption food values were cleaned. The cleaning procedure decided upon a maximum on per capita kg consumption of that food and any value above that was converted to the median of the non-missing values. The scheme for this is laid out in the file ***Household consumption data cleaning to median.xls.*** For each food it includes the maximum value permitted in terms of kilograms per person per day, and the number of cases changed to the median value with this condition. * in the first column indicates no limit maximum on per capita but they consumption was imposed and therefore no cases changed. The choice of values was based upon the histogram of the values reported, and a decision of those values that were to be considered as outliers because they represented a long and non-smooth tail to the distribution.

The mean age and sex adjusted household caloric requirement was calculated using the age and sex information collected in Section B-Household register and the basic energy requirement values from WHOs ***The management of nutrition in major emergencies.*** Geneva, 2000. The values used are presented in the accompanying table. The individual values for each household member are included as variable **AGESEXCALO** in ***NRVA 2003 HH Member data Register.dbf.*** These data were averaged to each household, **Mean_Age_Sex_calories**, and data **Total_cal_percapday**, is the total calories available to each household member per day. Therefore, if **Total_cal_percapday < Mean_Age_Sex_calorie;** household = calorie deficient.

In the dietary diversity section, a number of derived rise variables are included aside from the number of kilograms of each of the 64 food items included in the data set. These are:

Variable name	Interpretation
Food_item_KG	Amount in kgs consumed by the household during the last seven days
Food_item_pccal	Per capita per day kilocalories derived from each food item
Food_item_pc%	% energy derived from each food per capita per person per day

---

8 Household member data

Household member data is available in two forms; a series of DBF database format files for each of the 5 tables for household member data compressed together in one zip, ***NRVA 2003 HH Member data.zip***, containing:
NRVA 2003 HH Member data Register.dbf
NRVA 2003 HH Member data Migration.dbf
NRVA 2003 HH Member data Education.dbf
NRVA 2003 HH Member data Health.dbf
NRVA 2003 HH Member data Employment.dbf

Or an Access database:
NRVA Household data Access Database.zip

Data on household members was collected in the registry, migration, education, health and employment sections. All of the sections were in separate tables on separate pages of the questionnaire. As names were not employed to identify household members (not universally acceptable to Afghan culture), unique identification can only be made

with relationship to head of households and age together, available only in the registry table. The subsequent tables in the included in relationship to head of household and did not repeat the age of each individual.

This is ensure the linking of rows of data among these five tables can not be unambiguous completed for those household codes for which there are more than one code per household. For example, household head can be tracked successfully in all but 61 cases. In the case common multiple household codes, such as sons (8) and daughters (7), the situation is very different and tracking across rows is not reliably possible. Therefore, it is impossible to be sure that the enumerator has maintained the consistency of the row from one table to another. Where the data has the been recorded one the questionnaire on the same order across the five tables, there is the reasonable presumption that rows are matching, as was the instruction to the enumerators. Unfortunately, this was not adhered to in a significant number of cases, resulting in the order up of labelling of household individuals varying from table to table within the same household. In these cases, there is no unambiguous way of ensuring the household member data on a row in one table will match up with the same household member information on a row in another table if there is more than one household member with the same relationship ID. This was a design flaw in the questionnaire. Therefore the data for the five sections has been presented separately in separate DBF files to reinforce this issue.

From the table below, there are clearly some codes where almost all are occurring just once per household. This table gives the analyst and ability to judge which household codes can be reasonably tracked throughout the household member data, and which household codes against me more problematic for, for instance sons daughters and to a lesser degree brothers have a particular high frequency of multiple occurrences within households.

Household relation code	# per household	Count	Household relation code	# per household	Count
1	1	9631	9	1	778
1	2	60	9	2	213
1	3	1	9	3	58
2	1	1367	9	4	10
2	2	10	9	5	2
3	1	829	9	6	1
3	2	1	9	7	1
3	3	2	10	1	595
4	1	9743	10	2	127
4	2	346	10	3	39
4	3	16	10	4	7
4	4	2	10	5	3
5	1	2495	11	1	84
5	2	13	11	2	4
6	1	1470	12	1	53
6	2	5	12	3	1
6	3	1	13	1	60
7	1	2031	13	2	20
7	2	4076	13	3	5
7	3	3182	13	4	2
7	4	1212	14	1	54
7	5	355	14	2	22
7	6	89	14	3	3
7	7	28	14	4	4
7	8	11	14	5	1
7	9	3	15	1	83
7	10	1	15	2	49
8	1	1956	15	3	17
8	2	4073	15	4	5
8	3	3227	15	5	2
8	4	1345	16	1	222
8	5	399	16	2	66
8	6	115	16	3	25
8	7	29	16	4	8
8	8	15	16	5	3
8	9	3	16	6	2
8	10	3	17	1	5
8	11	1	17	2	1

9 Stopping rules

In the household questionnaire, not all stopping rules were obeyed. No attempt has been made to alter the data that has been entered after a stopping rule should have prevented further responses. May be a later date to release of this household data will include a modified data set. In the meantime, it is recommended that analysts evaluate each question and decide whether the information offered after enumerators have mistakenly continued after a stopping rule is appropriate to consider or not. Reporting on analysis should include the decisions analysts have made in this regard.