WVS 2011
Sampling Frame, Turkey
Yilmaz Esmer

<u>Target Population</u>: Individuals 18 years of age or above, who are residents of private households regardless of language, citizenship or legal status.

<u>Sampling Design</u>: Multi-stage full probability sampling

<u>Stage 1: Selection of PSU's (blocs)</u>
Turkish Statistics Institute has a "blocking" system whereby the whole population is divided into blocks of specified number of household addresses. The Institute provides randomly selected blocks with the specified number of addresses in each bloc.

For WVS Turkey 2011, 134 blocs (PSU's) with 300 household addresses in each bloc have been randomly selected. This number (i.e. 134 PSU's) will satisfy even the most rigid methodological standards.

<u>Stage 2: Selection of Households</u>
In each bloc, a fixed number (12) of addresses (households) are selected randomly from the list of 300 addresses.
Replacement: When it proves to be impossible to complete an interview in any given randomly selected address (refusal, address not found, demolished building, etc.) a randomly selected replacement address within the same bloc will be provided to the interviewer.
<u>Stage 3: Selection of Individuals</u>
Within the household, the individual to be interviewed will be selected with a Kish grid. Quota sampling is not allowed at all.

<u>Sample Size:</u>
With 134 PSU's and 12 interviews in each PSU, the theoretical sample size is 134x12=1608. Needless to say, some attrition is inevitable and the actual (realized) sample size will be somewhat lower than the theoretical size.

<u>Design Effect and Effective Sample Size</u>
The actual design effect can, of course, be calculated only after the full data set is available and is different for each question. Thus, the following is based on the assumption that the intraclass correlation within the PSU's is 0.01. The past experience has shown that this may be a realistic assumption.

The design effect (actual sampling error compared to SRS error) due to clustering is given by:
   1+ rho(nbar-1) where rho is the intraclass correlation and nbar is the average cluster size; in our case 12. Under the assumption that intraclass correlation is 0.01:
   Design effect= 1+0.01(12-1) =1.11 (again, this may be too optimistic for some
                                    questions and unduly pessimistic for others)
    Given this assumption:
   Effective sample size (i.e. the SRS equivalent)=1,608/1.11 = 1,448 minus some
                                    attrition
<u>Sample Representativeness</u>

Like all European countries, Turkey is also divided into NUTS (Eurostat's "Nomenclature of Territorial Units for Statistics")regions. At the NUTS-1 level, Turkey is divided into 12 regions. The distribution of the 134 PSU's in the sampling design among NUTS-1 regions is given below:

NUTS=1 region	# of PSU's	proportion of PSU's	adult pop 2010 census	proportion of pop/	difference PSU prop. minus pop proportion
1	24	0.1791	9,058,699	0.1865	-0.0074
2	7	0.0522	2,341,446	0.0482	0.0040
3	21	0.1567	6,955,532	0.1432	0.0135
4	11	0.0821	4,758,677	0.0980	-0.0159
5	15	0.1119	4,794,904	0.0987	0.0132
6	17	0.1269	6,118,634	0.1260	0.0009
7	9	0.0672	2,512,378	0.0517	0.0154
8	6	0.0448	3,154,425	0.0649	-0.0202
9	6	0.0448	1,739,704	0.0358	0.0090
10	4	0.0299	1,262,847	0.0260	0.0038
11	4	0.0299	2,013,566	0.0415	-0.0116
12	10	0.0746	3,856,287	0.0794	-0.0048
	134	1.0000	48,567,099	1	0.0000

As can be seen from the above table, the random selection procedure has resulted in an almost perfect distribution of PSU's for the whole country. Therefore, the representativeness of the initial sampling design is at the highest possible level. However, there will inevitably be some deviations due to variations in response rates and other unexpected outcomes. For Turkey, response rates are always higher in rural compared to urban units.