

Appendix 1: attrition and addition of new individuals: DHS, 1993 – 2005

Year of first appearance in panel	WAVE 1993	WAVE 1994	WAVE 1995	WAVE 1996	WAVE 1997	WAVE 1998	WAVE 1999	WAVE 2000	WAVE 2001	WAVE 2002	WAVE 2003	WAVE 2004	WAVE 2005
1993	5252 <i>100</i>	4475 <i>57</i>	3344 <i>45</i>	2503 <i>36</i>	1811 <i>29</i>	1068 <i>24</i>	843 <i>21</i>	541 <i>13</i>	526 <i>10</i>	452 <i>9</i>	375 <i>8</i>	366 <i>8</i>	359 <i>7</i>
1994		3442 <i>43</i>	2480 <i>33</i>	1517 <i>22</i>	1076 <i>18</i>	698 <i>15</i>	512 <i>13</i>	290 <i>7</i>	255 <i>5</i>	220 <i>4</i>	182 <i>4</i>	182 <i>4</i>	175 <i>3</i>
1995			1681 <i>22</i>	1018 <i>15</i>	704 <i>11</i>	469 <i>10</i>	312 <i>8</i>	133 <i>3</i>	151 <i>3</i>	125 <i>3</i>	107 <i>2</i>	99 <i>2</i>	100 <i>2</i>
1996				1862 <i>27</i>	1195 <i>19</i>	755 <i>17</i>	472 <i>12</i>	255 <i>6</i>	223 <i>4</i>	204 <i>4</i>	168 <i>4</i>	159 <i>3</i>	147 <i>3</i>
1997					1359 <i>22</i>	927 <i>20</i>	544 <i>14</i>	192 <i>5</i>	166 <i>3</i>	156 <i>3</i>	127 <i>3</i>	128 <i>3</i>	118 <i>2</i>
1998						620 <i>14</i>	428 <i>11</i>	208 <i>5</i>	192 <i>4</i>	163 <i>3</i>	152 <i>3</i>	143 <i>3</i>	141 <i>3</i>
1999							865 <i>22</i>	364 <i>9</i>	331 <i>6</i>	275 <i>6</i>	205 <i>4</i>	179 <i>4</i>	179 <i>3</i>
2000								2243 <i>53</i>	1546 <i>30</i>	1106 <i>22</i>	879 <i>18</i>	807 <i>17</i>	737 <i>14</i>
2001									1821 <i>35</i>	1017 <i>21</i>	797 <i>17</i>	712 <i>15</i>	645 <i>12</i>
2002										1237 <i>25</i>	825 <i>17</i>	676 <i>15</i>	587 <i>11</i>
2003											955 <i>20</i>	690 <i>15</i>	570 <i>11</i>
2004												486 <i>11</i>	377 <i>7</i>
2005													1082 <i>21</i>
Total	5252 <i>100</i>	7917 <i>100</i>	7505 <i>100</i>	6900 <i>100</i>	6145 <i>100</i>	4537 <i>100</i>	3976 <i>100</i>	4226 <i>100</i>	5211 <i>100</i>	4955 <i>100</i>	4772 <i>100</i>	4627 <i>100</i>	5217 <i>100</i>

Table A1: Attrition and addition of new individuals, DHS, waves 1993 – 2005
(Column) percentages in italics.

The table above gives an overview of attrition and addition of individuals in the DHS, waves 1993 through 2005. The column totals do not match those reported in the manuals accompanying every wave. Presumably this is caused by data clean-up by CentER. The figures include all members of the households under study, also those below the age of 16. A member of a household is any individual living within this household.

Looking through columns, the reader is able to verify the composition in terms of old and new individuals in each wave (percentage in italics). Looking across rows, the counts indicate how many of the individuals added in a given wave, have remained in the survey for the following years.

Appendix 2: construction of analysis files, DHS, 1993 – 2005

After completing the interview stage, and checks, new waves are made available to researchers through the DHS website: <http://www.uvt.nl/centerdata/dhs/>. The data, bar the questions added by De Nederlandsche Bank, are split in six files, plus two additional files with aggregated data computed by CentERdata. These latter two files contain user friendly income, asset and liability variables computed using the information collected on the participating individuals. We have assembled our analysis file from this data in the following four steps, using the SPSS syntax feature. We have elected to use the English version of both the datasets as well as the documentation.

Step 1: aggregation of datasets

The eight datasets per wave are downloaded in the SPSS portable format, and need to be converted to SPSS Data Documents first. The datasets are combined by computing a unique individual identification number, using the household number and the individual's member number within her household. For each wave, there is a household information (hhi) file which contains demographic information on all members of the sampled households, including those under 16. This file provides the researcher with a complete list of participants in each wave and is therefore best suited as the first file for the merger process. However, for the first wave, 1993, this file is missing. CentERdata advises to use the hhi file for 1994 instead and we have proceeded accordingly. Next to the thirteen completed wave files, we have constructed a single file with a list of all participating individuals throughout the years. This enables us to check attrition as has been done in Appendix 1, as well as the number of participations by individuals, as can be seen in table A2.

Number of participations	Frequency	Percent	Cumulative Percent
1	7324	31,98	31,98
2	4914	21,45	53,43
3	3293	14,38	67,81
4	2322	10,14	77,94
5	1882	8,22	86,16
6	1277	5,58	91,74
7	583	2,55	94,28
8	238	1,04	95,32
9	244	1,07	96,39
10	207	0,90	97,29
11	142	0,62	97,91
12	191	0,83	98,74
13	288	1,26	100,00
		100,00	

Table A2: Count of participations by individuals in the DHS, 1993 - 2005

The average number of times an individual participates in the DHS is 3.1; comparison of tables A1 and A2 reveals that some individuals have skipped certain waves. For example, from table A1 we

learn that 359 individuals participated in 2005, whom have entered the panel in 1993, implying 13 participations. However, from table A2, we learn that only 288 individuals actually did participate 13 times.

Step 2: data clean up

Although the definitions and coding of variables, labels and such has been remarkably stable and well documented for the DHS, inevitably changes have occurred. Wave by wave the data has been checked for resulting inconsistencies. All monetary amounts have been converted in euros. Ages were computed using year of interview and year of birth. Binary dummies are usually coded '1' and '2' (for example labelled 'yes' and 'no' respectively), such dummies have been recoded to '1' and '0' in order to facilitate use in multivariate models. The same holds for ordered categorical variables, for example the subjective evaluation of the household's financial position ranging from 'very bad' to 'very good': such variables have been recoded using a contrast coding scheme (i.e. '-1, 0, 1'). Unordered categorical variables, such as labour market status, were recoded into variables with combined categories of the original variable if needed, and into binary dummies for each (remaining) category, for interpretational ease.

Subsequent checks revealed a coding problem with the variable 'bezigbel' for the waves 2000 and 2001. Frequencies had been assigned the wrong labels, we have been provided with the following correction syntax by CentER:

```
RECODE bezigbel (1=1) (3=2) (4=3) (5=4) (6=5) (7=6) (8=7) (9=8) (10=9) (11=10) (12=12)
(13=13) (ELSE=COPY).
```

Checks on the corrected labour market status data revealed an additional problem: reported relative frequencies appear to suffer from a breach in trend from 2000 onwards as can be seen for the category 'employed' in figure A1 below. The variable appears to suffer from a change in interpretation by respondents, who were faced by new questionnaire routing in 2000 and onwards. Correcting for this is virtually impossible. Instead, we will use an alternative labour market status variable, referred to as 'bezig', which was answered by as much as 30% less respondents, because of routing.

One final important modification was made: all variables deemed useful for further analysis are assigned with a name and labels that reflect the year in which they were collected. When merging the various waves together in a later stage, this prevents problems with identification. Useful variables were saved in smaller, wave specific, files. These files included the variable 'account': this variable identifies the household member that has filled out the housing questionnaire.

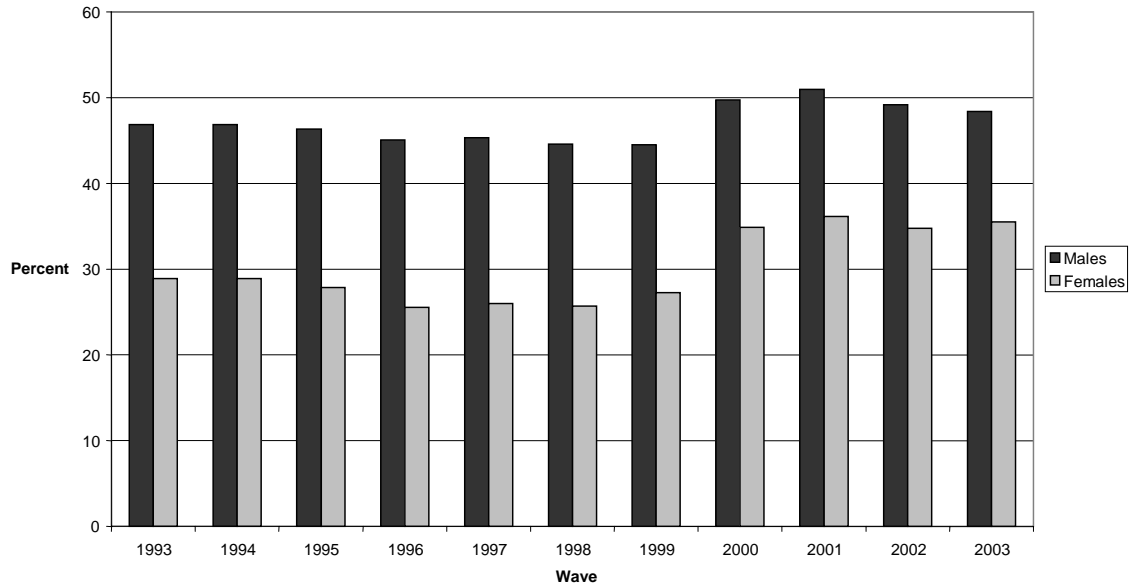


Figure A1: Relative frequencies for 'employed', corrected variable 'bezigbel', DHS 1993 – 2003.

Step 3: construction of the household file

Each record in the DHS is an individual, whereas in our study, the household is the unit of analysis. As each member of the household is identifiable through her member number, merging the files to obtain the household files appears straightforward. We have to keep in mind however, that only one member of the household was required to fill out the information that was valid for the household as a whole, for example the information with respect to housing. We found that the best way to identify this person was the use of the binary dummy variable 'account', which identifies the member of the household responsible for financial affairs. These members were then routed through the housing questionnaire. This was by no means always the head of the household, identified by the variable 'positie'. Also, throughout the years, the specific task of filling out the housing questionnaire shifted from one member to another on many occasions.

We proceeded by defining, for each household in each wave, the member for which account was '1' as the unit of analysis, and adding the information on other household members to this record. This procedure kept us from having to copy the housing variables for each member in each household, resulting in large files with many empty cells. We found that in many households, for more than one person account was '1', causing the programmed merging process to fail repeatedly. All these cases have had to be corrected by hand. This left us with thirteen household files, one for each wave.

The next step involved identifying husband and wife¹ for each household record, as ‘account=1’ had often shifted from one person to the next, member numbers and household position had no a priori relationship anymore. Using hierarchical syntax, data on gender and ‘positie’, we re-identified husband and wife from the data, and assigned the relevant demographic and economic variables. Among those was ‘age at time of the survey’: we decided to restrict ourselves to households with a wife present aged between 23 and 38, at the time of the survey. Refer to chapter 5 for detail on this matter. All other households were removed from the analysis. Finally, all resulting files were merged into our household file. This file contains a record for each household with a wife present aged 23 – 38 at any point during 1993 – 2005. For each household, information from the separate waves is stored in the columns, as well as averages across all observations. For those households experiencing first birth (see step 4), we used information on the household for the relevant points in time, to construct new data allowing us to ‘follow’ the household in the period surrounding first birth, as has been done in chapter 6.

Step 4: identification of first births

A seemingly straightforward step is the identification of those households that experienced first birth, while in observation. It is for these births that we can fully exploit the panel nature of our dataset, recording information at the time of birth, as well as before and after. The dataset contains demographic information on all children, whether they are present in the household at the time of survey (and hence appear as a separate record in the hhi file), or whether they live outside the household. The latter group does not appear as separate records, but rather as a group of variables in the record of one or more of their parents.

One crucial point of course is that the oldest child present in the household, is not necessarily the first born to that household: older siblings may live outside, or are missing from the household for other reasons. A child born during observation is only defined a first birth, when we have ascertained that there are no older children living inside or outside the household. Similarly, a household is not at risk of giving first birth, solely because they report that no children live with them at present.

We first recorded for each household, in each wave, what was the year of birth of the oldest child living in the household at the time of survey. If, for any wave, the year of birth for a child in the household was either equal to, or within 1 year of said wave, we assume the child to have been born in observation. As most variables are measured on a yearly basis, we can not achieve higher levels of precision than these two year spans. We then determined whether this child was in fact the oldest child living in the household at any point during the survey. If this was the case, we compared the year of

¹ Wife is meant to be married as well as cohabiting partner.

birth for these observed births to the year of birth of the oldest child living outside the household, if applicable.

Year of birth of the first child	Frequency	% of all households	Valid percent	Cumulative Percent
1992	33	1,0	8,3	8,3
1993	37	1,1	9,3	17,5
1994	45	1,4	11,3	28,8
1995	29	0,9	7,3	36,1
1996	18	0,5	4,5	40,6
1997	24	0,7	6,0	46,6
1998	14	0,4	3,5	50,1
1999	31	0,9	7,8	57,9
2000	37	1,1	9,3	67,2
2001	39	1,2	9,8	76,9
2002	25	0,8	6,3	83,2
2003	17	0,5	4,3	87,5
2004	31	0,9	7,8	95,2
2005	19	0,6	4,8	100,0
Total	399	12,2	100,0	
System missing	2879	87,8		
	3278	100,0		

Table A3: frequency of observed births, DHS, 1993 - 2005

Table A3 shows we observe 399 first births in 3278 households, with a wife present aged 23 – 38, during the period of 1993 to 2005.