Response to
"Some comments on the data used in Riphahn et al. (2003)"
by Jim Oeppen

July 18, 2013

Your point 3: You suggest that male hospital visits in counts 3 and 4-9 may be inverted.

Reponse: The paper lists 0.2 and 0.1 for 3 and 4-9 male hospital visits, respectively. In my old log-files I found these numbers

```
tab hospvis2 if female == 0

   hospvis2 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      13133       92.21       92.21
          1 |        880        6.18       98.39
          2 |        155        1.09       99.47
          3 |         21        0.15       99.62
          4 |         13        0.09       99.71
          5 |          7        0.05       99.76
          6 |          4        0.03       99.79
          7 |          4        0.03       99.82
          9 |          1        0.01       99.82
         10 |         25        0.18      100.00
------------+-----------------------------------
      Total |      14243      100.00
```

So we had a share of 0.15 percent observations for 3 and a total of 29/14243 = 0.203 percent for 4-9 hospital visits for men. The figures reported in Table I may be rounded for the 3 visits number, but they are clearly wrong for 4-9 visits. You are right.

Your point 4: The means for the year 1984 cannot be checked.

Response: For all years but 1984 we coded year indicators, 1984 is the reference group. If you consider the observations with all available year indicators coded zero that should describe the 1984 sample.

Your point 5.1: Within the paper Tables II and III are inconsistent.

Response: Unfortunately, this is true.

Your point 5.2: Some male variable labels are inverted.

Response: Unfortunately, this is true.

Your point 6.1: The 1987 handdum indicator appears to be inverted.

Response: I agree, your tables are clear on that. I checked my data coding and did not find an obvious error, i.e. the coding is identical across data years. I recall, however, that way back in past versions of the SOEP, the data had occasionally swapped indicator values for single variables in single years (i.e. they swapped e.g., zeros and ones just for one year). If this occurred for the handicap indicator, it would explain this particular problem. Clearly, the coding error should not have happened

Your point 6.1: … impossibly accurate decimal values.

Response: In all cases where you observe many decimal values in the data I had replaced missing values by gender - specific sample average values. By keeping the decimals the imputations are recognizable, one might drop these observations.

Your point 6.2: The 1987 handper values do not look plausible.

Response: I agree, this is connected to the handdum problem discussed above. My coding replaced missing values of handper based on pre-period values if the handdum indicator showed a handicap. So, when the handdum indicator was coded wrongly it generated miscodings of the handper variable.

Your point 6.3: The variables handdum and handper are inconsistent.

Response: I agree. This follows from the mistake in the 1987 handdum coding. If valid information was provided in the handper variable I never overwrote it. But when it should be there (based on handdum) and was missing I had used missing value imputations.

Your points 6.4 and 6.5: decimal responses on educ and hsat

Response: That is again the imputation for missing values.