

Some comments on the data used in Riphahn et al. (2003)

Jim Oeppen*

July 15, 2013

1 Introduction

In Table II of Riphahn *et al.* (2003) it is reported that about 20% of the respondents to the German Socioeconomic Panel (1984-1995) aged 25–65 are handicapped.¹ This is implausible and the following sections try to find the reason. I try to replicate the first three tables of the paper and then analyse some of the variables in their data. The results suggest major problems in the two variables related to handicapped status.

2 Data

The data were downloaded on July 10th 2013 from the Journal of Applied Econometrics website: <http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/>. My results below were calculated in R.

3 Table I, p.392

The numbers of observations by sex, and the percentages, for the downloaded data conform to the published Table I, p.39, except that Male hospital visits in counts 3 and 4-9 may be inverted.

*Research Scientist, Max Planck Institute for Demographic Research, Rostock, Germany.
email: oeppen@demogr.mpg.de

¹Regina T. Riphahn, Achim Wambach, and Andreas Million, 'Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation', *Journal of Applied Econometrics*, Vol. 18, No. 4, 2003, pp. 387-405.

4 Table II, p.393

The numbers of observations by sex, and the means, for the downloaded data conform to the published Table II, p. 393. The means for the year 1984 cannot be checked as they are not in the Table although they are included in the paper's analysis.

5 Table III, p.394

None of the female means in Table III can be reproduced exactly from the downloaded data, although the differences are relatively small. Footnote 7, p. 391, states that 'Observations of women who gave birth within the period of six months before and after a given calendar year were omitted from the sample for the relevant years.' This may explain why the table is inconsistent with the data. Within the paper, Tables II and III are inconsistent.

The male results are confirmed, except that the labels *yes* and *no* for the variables *Married* and *Children in household* are inverted. The data show that married men and men without children have higher levels of doctor and hospital visits than unmarried men and men with children: exactly the opposite of what is stated on pp. 392–3.

5.1 Greene (2009)

Greene (2009) also gives descriptive statistics for these data and reports that he downloaded the data from the same source.² The results of my descriptive statistics reported for Tables II and III are confirmed almost perfectly by Greene's Tables 1 and 2. The exception is the male mean for *addon*, where Greene's value in Table 2 is incompatible with the value for both sexes in his Table 1.

6 Problem Variables

6.1 handdum: handicapped = 1; otherwise = 0

Table 1 shows that almost 90% of adults aged 25–64 in 1987 are recorded as handicapped, compared to 8–14% in the other years. This is clearly impossible, and suggests that the coding may have been reversed. This is supported by cross-tabulation of *handdum* for individuals who are observed in both 1986 and 1987, and 1987 and 1988. Almost all individuals recorded with no handicap in 1986 and 1988 are recorded with a handicap in 1987.

In Table 1, and the variables in the tables below, some survey responses have impossibly accurate decimal values. The same values are repeated across different individuals.

²Greene, W. 'Models for count data with endogenous participation.' *Empirical Economics*, 2009, Volume 36, Issue 1, pp 133-173

Table 1: **handdum: handicapped = 1; otherwise = 0.**

observed values	frequency by year						
	1984	1985	1986	1987	1988	1991	1994
0	3420	3384	3391	408	3981	3863	3016
0.1581703	2	0	1	0	0	0	0
0.1687193	5	1	2	0	0	0	0
0.1943086	5	0	2	0	0	0	0
1	442	409	396	3258	502	477	361

6.2 handper: degree of handicap in percent (0 - 100)

As with *handdum*, the values in Table 2 for 1987 do not look plausible, as almost nobody reports zero on the handicap scale. The effect is to double the mean of *handper* in that year, relative to that observed in other years. The great majority of responses in 1987 are impossibly accurate decimal values, but all years show some decimal responses.

6.3 Covariance of handdum and handper

About 1% of individuals coded as zero in the handicap dummy variable *handdum* have percentage handicap greater than zero, the majority 50% or greater, which makes the two variables inconsistent. The distribution is shown in Table 3.

6.4 hsat: health satisfaction, coded 0 (low) - 10 (high)

Table 4: All years show some decimal responses.

6.5 educ: years of schooling

Table 5: All years show some decimal responses.

Table 2: **handper: degree of handicap in percent (0 - 100).**

observed values	frequency by year						
	1984	1985	1986	1987	1988	1991	1994
0	3420	3384	3391	27	3981	3863	3016
3.396562	2	0	2	761	9	2	2
6.701853	11	9	9	804	10	1	2
7.649712	8	4	6	1627	6	10	5
10	5	2	1	6	1	1	2
15	2	1	0	4	1	2	0
20	19	15	15	19	18	23	18
23	0	0	0	0	1	1	0
25	5	5	7	6	4	4	4
26	0	0	0	1	1	0	1
27	0	0	0	0	1	0	1
28	0	0	0	1	0	1	0
29	0	0	0	1	0	0	0
30	43	38	33	47	35	63	57
35	2	0	0	2	1	1	2
40	33	30	31	38	43	38	34
45	0	1	1	1	2	3	0
48	0	1	1	1	0	1	0
49	0	1	1	0	0	0	0
50	104	111	113	114	130	118	88
55	2	1	1	2	0	1	0
60	55	47	44	59	64	62	47
65	0	0	0	1	1	0	0
66	0	2	2	0	0	0	0
67	0	0	0	0	0	0	1
70	51	43	44	44	43	42	29
75	0	1	2	2	1	2	0
80	55	47	40	40	51	43	28
82	0	0	0	0	0	0	1
85	0	0	0	0	0	1	0
90	19	14	10	13	19	8	3
95	0	1	1	0	0	0	0
100	38	36	37	45	60	49	36
Mean Response	6.58	6.16	5.95	12.48	6.44	6.06	5.70

Table 3: **Individuals with zero handicap dummy but percent > 0.**

Percent Handicap	frequency
10	3
15	2
20	9
25	4
26	1
28	1
29	1
30	35
35	1
40	32
45	1
50	100
55	1
60	57
65	1
70	40
75	1
80	37
90	13
100	41
Total	381

Table 4: **hsat: health satisfaction, coded 0 (low) - 10 (high).**

observed values	frequency by year						
	1984	1985	1986	1987	1988	1991	1994
0	101	61	63	52	66	55	49
1	41	44	29	34	52	29	26
2	112	79	74	88	112	94	83
3	155	139	151	150	196	195	187
4	159	170	219	168	237	241	196
5	636	647	599	571	680	638	462
6	276	325	322	369	434	447	357
6.583997	0	1	1	0	0	1	3
6.877203	4	6	3	3	5	4	2
6.909199	0	1	1	2	1	1	1
7	502	562	550	577	713	710	577
8	749	797	849	866	1073	1041	797
9	393	425	441	383	500	537	382
10	746	537	490	403	414	347	255

Table 5: **educ: years of schooling.**

observed values	frequency by year						
	1984	1985	1986	1987	1988	1991	1994
7	18	16	17	19	24	38	40
8.5	3	4	4	3	4	12	24
9	881	783	745	710	762	665	488
10	82	80	82	85	111	116	85
10.5	1384	1337	1329	1293	1528	1417	1077
10.80549	3	4	6	7	5	6	6
11	378	389	376	351	435	410	302
11.44142	2	3	1	3	5	8	14
11.5	397	398	414	400	508	532	421
11.81824	4	6	6	9	19	15	16
12	203	212	224	220	294	313	262
13	64	69	79	80	116	113	90
13.5	22	24	23	22	37	36	49
14	41	47	47	44	56	58	41
14.5	49	56	53	48	68	82	73
15	81	93	95	100	126	136	92
16	32	39	36	33	53	55	39
17	9	11	12	9	9	13	11
18	221	223	243	230	323	315	247