Schmollers Jahrbuch 127 (2007), 227 – 258
Duncker & Humblot, Berlin

Comparing Economic Outcomes of Populations with Disabilities

A Method for Comparing the Economic Outcomes of the Working-Age Population with Disabilities in Germany and the United States

By Richard V. Burkhauser and Mathis Schroeder*

Abstract

Using a work limitation-based measure of disability, researchers have found that the employment of working-age men (aged 21 – 58) with disabilities fell dramatically relative to such men without disabilities in the United States in the 1990s. Because no such measure of the population with disabilities is consistently provided for Germany, this paper develops two alternative work limitation-based measures of disability in the German Socio-Economic Panel (SOEP). Using these measures we find that while the relative employment of working-age men in Germany fell in the 1980s, in contrast to the United States, it rose over the 1990s.

Zusammenfassung

Studien in den USA haben für die 90er-Jahre einen erheblichen Rückgang der Beschäftigung von behinderten Männern relativ zu Männern ohne Behinderung festgestellt. Die Definition dieser Behinderungen basiert auf einer Einschränkung der Arbeitsfähigkeit. Da eine vergleichbare Definition für Deutschland nicht fortlaufend zur Verfügung steht, entwickelt dieser Beitrag zwei mögliche Maße für Behinderung im Sozio-Oekonomischen Panel (SOEP). Anhand dieser Maße stellen wir fest, dass zwar die relative Beschäftigung von behinderten Männern in Deutschland in den 80er-Jahren gefallen ist, sie aber im Gegensatz zu den USA in den 90ern wieder anstieg.

JEL Classifications: C81, F02, J80

Received: 05. 06. 2006
Accepted: 13. 02. 2007

* Partial funding for this work came from the U.S. Department of Education, National Institute of Disability and Rehabilitation Research (NIDRR), cooperation agreements H133BO40013 and H133BO31111. In addition, Schroeder received a pre-doctoral stipend from the Steven H. Sandell Grant Program for Junior Scholars in Retirement Research, funded by a grant from the United States Social Security Administration (SSA) through the Michigan Retirement Research Center. We thank two anonymous referees, J. S. Butler, Dean R. Lillard, Ludmila Rovba and Robert R. Weathers II for helpful comments and suggestions. All remaining errors are our own.

1. Introduction

Most OECD countries regularly survey a large representative sample of their population to document economic well-being and labor market outcomes. They do so to monitor how their general population is doing over time as well as to see how more economically vulnerable subgroups are faring – e.g. children, women, older people, or racial and ethnic minorities.

In most cases, these potentially more vulnerable populations can be reasonably captured by a single variable – e.g. age, sex, race, or geographic location. But in some cases, it is more difficult to do so, either because of ambiguities in the conceptualization of the population or difficulties in capturing this population with a single question or even a series of questions, or both. In other cases, it is because the questions used to capture the population have evolved over time and it has been difficult to create *ex post* rules that permit comparisons across the data. All of these issues confront researchers interested in comparing the economic outcomes of working-age people with disabilities in Germany and the United States. This is the case even though substantial efforts have already been made to produce comparable German and United States socio-economic data sets, such as the Cross-National Equivalent File (for a detailed discussion of CNEF see: Burkhauser / Butrica / Daly / Lillard, 2001).

The population of working-age people with disabilities in social sciences-based nationally representative datasets is usually determined by the response to a single work limitation question. Researchers have used data from this type of question to show that the employment of working-age men (aged 21 – 58) with work limitation-based disabilities fell over the 1990s relative to their counterparts without disabilities.[1] (See for example: Acemoglu / Angrist, 2001; Autor / Duggan, 2003; Bound / Waidmann, 2002; Hotchkiss, 2003; and Burkhauser / Houtenville / Rovba, 2006.)

In large part because of a lack of equivalent data, few cross-national comparisons of the employment of those with disabilities have been done in this literature. This paper attempts to overcome the lack of equivalent international disability data by creating a consistently estimated work limitation-based measure of disability for Germany using data from the German Socio-Economic Panel (SOEP). While the SOEP has asked a work limitation question in some of its waves, it has not asked it in all waves and is unlikely to ask it in the future. As a result, researchers have relied on health satisfaction measures or have only used years of the SOEP that included the work limitation question. For example, Riphahn (1999) investigates the impact of health shocks on em-

[1] The choice of ages that make up the working-age population is somewhat arbitrary. We chose those aged 21 – 58 because this is the age range used by Acemoglu and Angrist (2001). More generally we do so to abstract from differences in employment rates related to entry and exit from the labor market of those without disabilities.

ployment and economic well-being of older German workers. Because the SOEP does not ask this question each year, she uses changes in health satisfaction to define a health shock, thereby abstracting from relevant information about the individual's health related to work. Burkhauser and Daly (1998), on the other hand, use the work limitation question as their measure of disability, but have to limit their study to the years over which this question was asked.

A work limitation-based disability variable for all SOEP years is therefore desirable both for cross-national comparative research as well as for dynamic analyses of the population with disabilities within Germany.[2] We create two work limitation-based measures of disability in the SOEP that are comparable to the one based on the work limitation question in the CPS and other United States datasets. These measures are alternatives to other SOEP measures of health limitation and have the comparative advantage of being consistently defined and available over all waves of the data. We evaluate the measures in detail based on those years a work limitation question is available and also provide an example of the power of such a measure. We show that the dramatic decline in the employment of working-age men with disabilities relative to such men without disabilities that occurred in the United States in the 1990s occurred in the 1980s in Germany. The relative employment of working-age men in Germany rose during the 1990s.

2. The Problem

The SOEP is a longitudinal survey of German residents administered by the German Institute for Economic Research (DIW). It started in 1984 in the western German states with a representative sample of over 12,000 individuals in almost 6,000 German and guest worker households. Several changes in sampling have occurred. With the unification of Germany more than 2,000 households in the eastern states of Germany were added in 1990, and in 1994 and 1995 new samples of immigrants (around 1,000 individuals) were included. In 1998, a refresher sample added around 1,000 households. In 2000, the innovation sample almost doubled the sample size of the existing SOEP, with more than 10,000 new respondents, resulting in a total of around 24,000. We use the SOEP 100 percent sample for the years 1984 – 2002. These data are distributed directly to researchers by the DIW.[3]

---

[2] The Health and Retirement Survey (HRS) in the United States and the Survey of Health, Aging and Retirement in Europe (SHARE) have measures of disability that are directly comparable. Compared to the SOEP, however, these surveys are limited in the type of respondents that are asked (age 50 and older) and in the duration of their panels: While the HRS started in 1992, the SHARE only started in 2004.

[3] For more information, see: Wagner / Burkhauser / Behringer (1993) and http://www.diw.de / english / sop / index.html (accessed September 2006).

The SOEP includes several questions related to work limitations, which vary by year and by content. Because we want to find a measure of disability that is comparable to the work limitation-based measure used in United States datasets, we focus on questions in the SOEP that provide similar information.[4] In the United States Current Population Survey (CPS) the respondent is required to answer the following question for every working-age person in the household: *"Does anyone in this household have a health problem or disability which prevents them from working or which limits the kind or amount of work they can do?"* The question asked of the head in the Panel Study of Income Dynamics (PSID) is: *"Do you (Head) have a physical or nervous condition that limits the type of work or the amount of work you can do?"* From 1981 onward the PSID asks the respondent to answer this question for the household head and spouse.[5] In the SOEP, the following question (reported for the sample older than fifteen) is closest to the questions in the CPS and the PSID: *"Disregarding occasional illnesses, does your health limit your daily activities in the household, on the job or in school, and to what extent?"*[6]

The CPS only asks for a *"yes"* or *"no"* response to its disability question. The PSID has a follow-up question for those who say *"yes"* that asks them about the degree of the limitation on their activities and allows them to answer: *"a lot," "somewhat,"* or *"just a little."* The SOEP directly asks about the severity of a possible disability, with categories *"Not at all," "slightly,"* or *"severely."*[7]

---

4 Even though the wording of the questions is very similar, cultural differences between the United States and Germany (e.g. stigmas associated with disability) could lead to differences in reported levels and trends for disabilities. For a discussion of these issues see: Kapteyn / Smith / van Soest (2006).

5 For further information on the health and disability information in the PSID, see Andreski et al. (2005). For an evaluation of the value of the PSID for research on the working-age population with disabilities, see Burkhauser / Weathers / Schroeder (2006).

6 The exact question is: "Von kurzen Erkrankungen einmal abgesehen, behindert Sie Ihr Gesundheitszustand bei der Erfüllung alltäglicher Aufgaben, z. B. Haushalt, Beruf oder Ausbildung? In welchem Umfang?" (SOEP, 2000). This question captures a somewhat broader spectrum of activities, since it entails household and school as well. Hence, compared to the PSID or the CPS, ceteris paribus, more people are likely to report such an activity limitation.

7 It is unclear, to what extent the possibility of having three levels of disability in the SOEP influences the respondents' answers. Assuming that the underlying distribution of disability is continuous, allowing for three instead of two levels of disability could shift the thresholds at which individuals consider themselves to be disabled. This shift occurs if the group reporting a slight limitation consists of individuals from the healthy as well as the disabled sample when (hypothetically) faced with the choice of reporting any disability or none. This problem is empirically addressed by Banks / Kapteyn / Smith / van Soest (2004), who specifically test the influence of moving from a five-point disability scale to a two-point scale. The five possible answers are "not limited",

Table 1 considers issues related to differences in the prevalence of work limitation-based disabilities across the CPS, PSID and SOEP. In all years, the CPS working-age population (aged 21 – 58), has a much smaller prevalence of work limitation-based disability (column 1) than does the PSID working-age population (column 2). This difference is significant, but may be attributed to the slightly different CPS sample (all members of the household, not only heads and spouses), the different questioning schemes, and the general difference between a cross-sectional (CPS) and a longitudinal panel (PSID) rather than to true difference in the prevalence of work based disability. But it may also be due to the multiple layers of severity offered to PSID respondents. For instance, if those who report that their work limitation bothers them only "a little" in the PSID are excluded from the work limitation-based population (column 6), the differences in the prevalence in the two data sets are dramatically reduced (see also discussion in footnote 7).

Comparisons between the CPS and the SOEP have similar problems with the addition of potential cultural differences in how workers in Germany respond to such questions relative to the United States. As column 7 shows, the unconditional prevalence rates of work limitation-based disability of the German working-age population is far higher than the prevalence rates reported in either the CPS or PSID. This is similar to findings for the Netherlands by Banks / Kapteyn / Smith / van Soest (2004). In part, this difference may also be due to the larger scope of the SOEP question, which, as discussed above, includes school and household as activity limitations.

However the fraction reporting a severe work-limitation-based disability (column 9) is much smaller and is much closer to the prevalence rates reported in the CPS and PSID. Hence, using the United States datasets as a reference point, we define the population reporting a severe limitation in the SOEP as our population of interest. In constructing our disability measures, we include individuals regardless of age since we are primarily interested in a good prediction and thus take advantage of the larger sample size.[8]

"some-what limited", "rather limited", "severely limited", and "very severely limited". The two-point scale is simply "yes" and "no." Their analysis concentrates on the percentage in the 5-point scale categories that report a "yes" when faced with the two-point scale. For category 1, 4.3 % report a disability. For category 2 ("somewhat limited"), they find that about 56 % report a disability on the two-point scale. For the categories 3 – 5, at least 91 % report a disability, which suggests that while there is some underreporting due to the two-point scale, it is mainly a problem for those on the margin of a disability. While this implies that of the "somewhat limited" category in the SOEP we are losing about half (10 % of the total population) that would classify themselves into being disabled, it seems very likely that those reporting a severe limitation would have reported a limitation if it was on a 2-point scale.

[8] Calculations with a sample of working-age individuals (21 – 58) lead to worse predictions as measured by the methods introduced in section 3. Specifically, the predictions overestimate the population with disabilities by more than 30 percent.

Table 1: **Prevalence of Work Limitations in CPS, PSID and SOEP by Year[a)]**

Year	CPS	PSID[b)]					SOEP[c)]		
	Any[d)]	Any[d), e)]	A Little	Somewhat	A Lot	More Than a Little[f)]	Any[d)]	Slightly Limited	Severely Limited
1984	0.062	0.112	0.040	0.035	0.037	0.071	0.288	0.205	0.083
1985	0.065	0.120	0.042	0.036	0.041	0.077	0.298	0.223	0.074
1986	0.064	0.102	0.037	0.032	0.020	0.064	0.312	0.236	0.076
1987	0.065	0.125	0.042	0.031	0.031	0.082	0.304	0.240	0.064
1988	0.062	0.135	0.045	0.037	0.030	0.090			
1989	0.062	0.127	0.037	0.037	0.031	0.090			
1990	0.063	0.138	0.041	0.041	0.031	0.096			
1991	0.064	0.133	0.039	0.037	0.031	0.094			
1992	0.066	0.121	0.042	0.034	0.026	0.079	0.286	0.227	0.059
1993	0.069	0.114	0.042	0.030	0.025	0.072			
1994	0.072	0.122	0.045	0.035	0.030	0.077			
1995	0.073	0.123	0.041	0.037	0.032	0.083	0.296	0.238	0.057
1996	0.072	0.137	0.044	0.039	0.035	0.093	0.295	0.236	0.059
1997	0.073	0.129	0.042	0.037	0.031	0.088	0.294	0.237	0.057
1998	0.071						0.279	0.222	0.058
1999	0.069	0.129	0.041	0.036	0.030	0.089	0.275	0.217	0.057
2000	0.070						0.262	0.209	0.052
2001	0.069	0.134	0.045	0.041	0.029	0.089	0.266	0.211	0.055

Notes:
a) All means are weighted by the respective individual weights and depict the fractions for working-age (21 – 58) individuals.
b) The PSID was not administered in 1998 and 2000.
c) The work limitations-based disability question was not asked in the SOEP in 1988 – 1991 and 1993 – 1994.
d) "Any" refers to a reported limitation in the CPS and the PSID and to the sum of those reporting severe and slight limitations in the SOEP.
e) The sum of "a little," "somewhat," and "a lot" in the PSID do not add up to the "any" fraction after 1985, since the filtering scheme was changed. Individuals who reported not being able to work at all were not asked the next question about the severity of the limitation.
f) "More than a little" refers to the percentage reporting "a little" limitation subtracted from the percentage reporting any limitation.

Source: Authors' calculations using CPS 1984 – 2001, PSID 1984 – 2001, and SOEP 1984 – 2001.

The SOEP limitation-based disability question is only asked in 12 of the 19 waves of data we use here.[9] It was not asked from 1988 to 1991 and in 1993, 1994, and after 2001. To bridge these gaps in the data we combine two other health-related SOEP questions available in all waves of the SOEP data to consistently construct alternative work limitation measures and test their performance by comparing them to the actual reported work limitation in the years these data are available.

3. Constructing a Measure of Work Limitation-Based Disability

While respondents are asked other questions related to their health in the SOEP, only current health satisfaction is available in all survey years. Since individuals self-reported work limitations are likely to be related to their perceptions of their health, and since health satisfaction is a completely subjective measure that is likely to be highly correlated with a self-perceived work limitation, this variable is one part of our measures of work limitation-based disability.

Another question in the SOEP that is directly related to work limitations and disability is: *"Are you officially registered as having a reduced capacity to work or as being severely disabled? If yes, what is the degree of your disability?"* Individuals register their disability with the German Pension Office, which also assigns a degree of disability, ranging from 1 to 100 percent (see Appendix 1 for a discussion of the German disability registration system). While this question was not asked of everyone in 1986, 1990, and 1993, the DIW provides an algorithm to impute this value.[10] We use this algorithm to obtain incidence and degree of a registered disability for all years, but do not use 1986, 1990 or 1993 in the construction of our measures of disability.

The registration-based disability question directly captures part of the population with disabilities. However, it is not likely to be directly comparable to the United States version of a self-reported work limitation question. This is the case since, in addition to having a disability, one must go through an arduous administrative process to become registered. If we only used this variable

[9] While SOEP data are available for more recent years, the questions about severe limitations are not asked after 2001 and hence the newer waves could not contribute to our analyses.

[10] The value in the previous year is assigned to the year the question is missing. See page 2–116 in the CNEF code-book: http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/upload/GSOEP_CNEF_2.pdf (accessed May 2007).

to define work limitations, we would miss individuals with a work limitation who were in the process of becoming registered, had short-term limitations, or decided it was not worth registering. For this reason, we create a set of alternative work limitation-based disability definitions by combining information on both registration and health satisfaction.

Table 2 reports the average yearly level of health satisfaction and the average degree of official disability for those who do and do not report a severe work limitation-based disability. (For simplicity, standard deviations are not reported in Table 2. However, all differences between these two sub-samples are significant at the 1 percent level.) The level of health satisfaction is measured on a scale from zero (not satisfied at all) to ten (most satisfied). The degree of official disability is measured on a scale from zero to one hundred, where zero indicates that the person is not registered and regarded as being in perfect health and one hundred is a registered individual with one severe disability (e.g. missing a limb) or several smaller limitations.

Those with severe work limitation-based disabilities have a mean level of satisfaction with health (column 1) which across all years is on average about one-half that reported by the rest of the population (column 3). In addition, they have a mean reported degree of disability (column 5) which is on average across all the years about ten times that of the rest of the sample. The correlation between health satisfaction and a severe limitation is –0.67, while the correlation between being registered as disabled and reporting a severe limitation is 0.72. These are both highly significant correlations. Thus we argue that some combination of the level of one's health satisfaction and the degree of one's registered disability will be a plausible alternative definition for a severe work limitation-based disability.

Having identified two variables that are highly correlated with a severe limitation, we offer two possible ways to construct a consistent measure of disability in all years. The first is to predict what we regard as the true disability status with an estimation of limited dependent variables, using the health satisfaction and the degree of disability as explanatory variables. The second is to find combinations of different levels of health satisfaction and degrees of disability, and see how these match with the true measure of limitation. While the estimation method may be preferred because it leads to a better use of the information and preserves the marginal distributions, it also leads to the problem of imprecise predictions if the standard errors are large. Although the single measure of combinations of health satisfaction and degrees of disability does not make maximum use of the data, it does not suffer from prediction biases as such and is somewhat easier to compute.

Table 2: **Average Yearly Level of Health Satisfaction and Degree of Disability of those Reporting or Not Reporting a Severe Work Limitation-based Disability[a)]**

Year	Average Yearly Level of:							
	Satisfaction with Current Health[b)]				Degree of Disability[c)]			
	Severely Limited	*N*	Not severely Limited	*N*	Severely Limited	*N*	Not severely Limited	*N*
1984	3.25	1468	7.42	10714	34.88	1464	2.84	10673
1985	3.48	1204	7.27	9802	33.99	1199	3.24	9778
1986[d)]								
1987	3.46	1036	7.15	9429	32.82	1032	3.51	9387
1992	3.03	1224	7.13	12092	35.90	1214	4.13	11983
1995	3.11	1068	6.84	12532	37.48	1059	4.54	12467
1996	3.24	1092	6.82	12320	37.48	1092	4.44	12293
1997	3.22	1028	6.79	12174	35.81	1021	4.59	12146
1998	3.55	1209	6.86	13317	39.87	1208	4.55	13259
1999	3.15	1164	6.86	12834	41.68	1169	4.59	12845
2000	3.24	2099	7.02	22366	39.58	2094	4.07	22348
2001	3.39	1953	7.03	20245	38.82	1950	4.20	20253
Correlations[e)]	**Satisfaction and Severely Limited**				**Registered and Severely Limited**			
Value (Standard Error)	–0.666 (0.002)				0.723 (0.003)			

Notes:
[a)] All calculations are weighted. *N* refers to the number of non-missing observations for the specific questions in a given year. Values for self-reported severe limitations were not collected over 1988 – 1991, in 1993, 1994, and 2002.
[b)] Scale from 0 – 10, 0: not satisfied at all, 10: completely satisfied.
[c)] Scale from 0 – 100, 0: not disabled, 50 or more: officially considered severely disabled.
[d)] 1986 is omitted since the degree of disability is only imputed in this wave (see text for details).
[e)] The correlations are estimated using the procedure *polychoric* in Stata™. See http: // www.unc.edu / ~skolenik / stata, accessed September 2006.

Source: Authors' calculations using SOEP 1984 – 2002.

To evaluate our predictions based on the true measure, we capture the possible outcomes by comparing the true and the constructed measures in a 2-by-2 matrix of weighted observations:

		Severely Limited	
		No	**Yes**
Disabled by Constructed Measure	**No**	*A*	*B*
	Yes	*C*	*D*

Each cell represents the number of observations or sample fractions, where cell *A* contains true negatives, cell *B* contains false negatives, cell *C* contains false positives, and cell *D* contains true positives. A variety of indices are used in biology to evaluate the similarity between two species, an approach we adopt here to evaluate the match between our predictions and the true measure. Any index evaluating the goodness of fit uses the cells of the above matrix in some form. The following index is the simplest one – it computes the percentage of matches in the total population:

$$S_B = (A + D)/(A + B + C + D) \ .$$

This function is maximized at $S_B = 1$, i.e. when *A* and *D* capture the whole population. However, this approach may not always be suitable to the researcher. Especially in our case, the low fraction of disabled in any year causes a problem: Consider the following two matrices to evaluate two hypothetical definitions *X* and *Y*, where 15 percent of the sample is severely limited and 85 percent is not:

		(a) Severe Limitation				(b) Severe Limitation	
		No	**Yes**			**No**	**Yes**
Disabled by Definition *X*	**No**	75	5	**Disabled by Definition *Y***	**No**	84	14
	Yes	10	10		**Yes**	1	1

The above index S_B leads to a value of $S_B = 0.85$ for both definitions ((75 + 10) / 100 for panel (a), (84 + 1) / 100 for panel (b)), although definition *Y* is less valuable for our purpose, since it defines far fewer individuals as severely limited.

One index that takes heterogeneity across the columns into account is the *Pearson's Phi,* first introduced by Pearson (1904):

$$S_\phi = (AD - BC)/\left[(A + B)(A + C)(D + B)(D + C)\right]^{1/2} \ .$$

S_ϕ is maximized at 1 when $B = C = 0$, and reaches its minimum at –1, when both A and D are zero. However, similar to a correlation coefficient, as negative values approach –1, the association becomes larger. The smallest association is at zero, when the product of true positives and negatives equals the product of false negatives and positives. (To avoid an undefined index, S_ϕ is set to 0 if any of the sums in the denominator are zero.) The respective products increase the index's sensitivity with respect to unequal columns – in the previous example, definition X would yield a value of $S_\phi = 0.49$, whereas definition Y would only yield a value of $S_\phi = 0.14$. (See Appendix 2 for a more detailed discussion on the differences between the two indices.) While the small fraction of people with a severe work limitation-based disability in our population justifies our using this index, this choice remains somewhat arbitrary.[11]

We now turn to the construction of our work limitation-based indicator variables: I^L, stemming from a logit estimation, and I^C, from a simple combination of health satisfaction and degree of registered disability. The goal of the logit approach is to estimate the likelihood of a disability (i.e. the indicator of a severe limitation is equal to one) using the health satisfaction and the degree of disability as explanatory variables.[12] Using the estimated influences, we can then predict the probability of having a disability. The indicator variable I^L is created from these probabilities by using the probability that maximizes the above index S_ϕ as the cutoff point, i.e. anyone above the cutoff probability is defined as having a disability.

The combination approach uses different levels of health satisfaction and of the degree of disability to create the indicator variable. We compute the value for index S_ϕ for all combinations of health satisfaction and the official degree of disability. Following the direction of the correlations in Table 2, we take the health satisfaction levels as upper bounds and the degree of disability as lower bounds, i.e. our constructed measure involves all individuals of a certain level of health satisfaction *or below* and a certain degree of disability *or above*. We allow for definitions where both these criterions have to be fulfilled (i.e. level of health *and* degree of disability) as well as for those where any of the two are met (level of health *or* degree of disability). We then compare all definitions created this way by computing the index value S_ϕ and chose the indicator I^C with the highest value.

---

11 An alternative method of finding the optimal measure would be to see how a constructed measure performs compared to the true measure in a regression using disability status as an explanatory variable. However, a different kind of arbitrariness is introduced here: the results depend strongly on the regression equation, meaning that different measures are closest to the true measure when changing the dependent variable and / or the other explanatory variables.

12 In principle, other explanatory variables could be added to the logit estimation. For example, age and gender could be important. However, it is not clear which variables should be used, and additionally, the comparison with the combination measures becomes less straightforward.

The estimation approach takes the given information into account better than the combination measure. However, the combination measure is easier to compute and, more importantly, is not subject to the choice of an estimation procedure or the imprecision of prediction. Hence even if the estimation measure yields better results and can be regarded as the upper limit in the efficient use of information available, we are still interested in how well the non-estimated measure will do in comparison.

4. Results

Table 3 shows the estimation results from the logit estimation using health satisfaction and the degree of disability to predict the severe limitation. The results match the findings in Table 2. Both variables have highly significant point estimates and contribute to the probability in the expected way. While an increase in health satisfaction is associated with a decrease in the likelihood of being severely limited, an increase in the degree of official disability increases the chances of reporting a severe limitation. The marginal effects evaluated at the means of those reporting a severe limitation suggest that for a ten percentage point increase in registered disability, the probability of reporting a severe limitation increases by almost one percent, whereas increasing the level of health satisfaction by one unit decreases this probability by 16 percent. Measured by the standard error, all effects are very precisely estimated. The predicted probability that maximizes S_ϕ is 29.11 percent. For comparison, the probability that would maximize the number of correct matches (and thus the index S_B) is 47 percent.[13]

The best concordance of a combination definition with the severe limitation (as measured by the index S_ϕ) is a combination of individuals with a health satisfaction level of at most 2 or a degree of disability of at least 53 percent. (Appendix Figure 2 plots the index for combinations of health satisfaction levels and the degree of disability for relevant ranges.)

As expected, the measure derived from the logit estimation does slightly better. The index value is $S_\phi^L = 0.576$ vs. $S_\phi^C = 0.558$ from the combinations of level of health satisfaction and the degree of disability. Both measures have a standard error of 0.002 (based on Pearson and Heron, 1913) so that the difference between the index values is statistically significant. We now turn to analyzing how the two measures differ in predicting the population with severe limitations, and if the differences are of relevance.

[13] The function we try to maximize (the value of S_ψ with respect to the cutoff probability) is not strictly monotonic. We find the maximum numerically by varying the cutoff probability and computing the respective values for the index. The degree of precision employed is 1 in 10,000. There are no changes in the samples if the precision is increased.

Table 3

Results From Logit Estimation: Predicting a Severe Limitation

	Coefficient	Marginal Effects
Degree of Disability	0.0314 (0.0008)	0.0078 (0.0002)
Satisfaction With Health	–0.6532 (0.0101)	–0.1619 (0.0027)
Constant	0.8049 (0.0502)	
Observations	159068	
Pseudo R^2	0.4178	

Notes: Dependent variable is the indicator of having reported a severe limitation. Weighted estimation based on the full sample and all years the variables are not missing. 1986 is excluded, see text for details. Standard errors (in parentheses) allow for correlations within a person over the years. Marginal effects are evaluated at the means of those reporting a severe limitation: Mean satisfaction with health is 3.3, the mean degree of disability is 37.2.

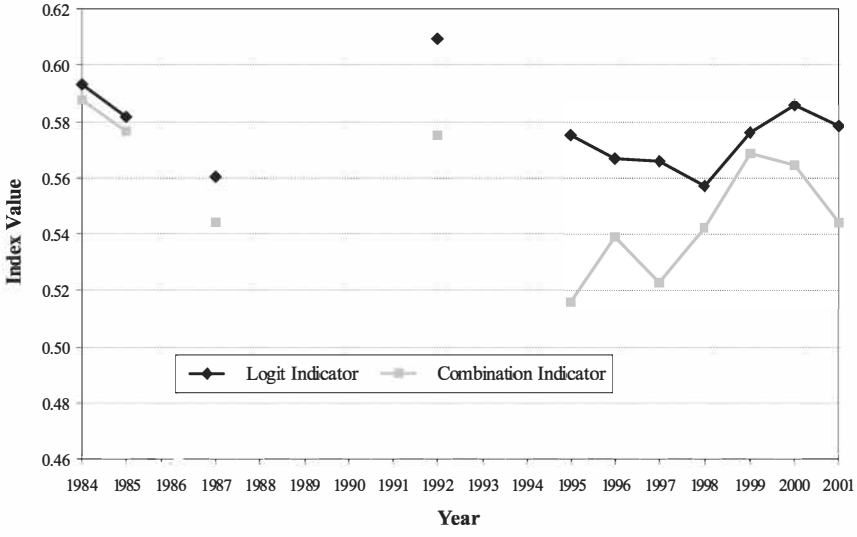
Source: Authors' calculations using SOEP 1984 – 2002.

Figure 1 shows the yearly index values for the two definitions, for those years the severe limitation is available and the satisfaction and degree of disability measures need not be imputed (note that the base is not zero). Similar to the index values for all years taken together, in each year the indicator from the logit estimation has a higher value than the combination indicator. While the overall difference between the two indicators is about four percent, some years have larger differences – especially in 1992 and between 1995 and 1997, the value for the combination indicator is between five and ten percent lower.

Table 4 shows how these differences in the index value translate into the actual samples. The first two rows show the match rates for all years as well as separately by year. Both measures consistently place more than 90 percent of the sample in the correct category. Although the difference between the indicators varies by year, it remains small. The largest difference in any year is one percentage point in 1995.

However, as discussed above, the overall percentage match is not the only factor we consider. Rows three and four show the prevalence of a severe limitation predicted by the two indicators; row five depicts the actual sample percentage of reported severe limitation. The logit predictions are somewhat closer to the true values, but both indicators perform fairly well over the years. This relationship is also shown in Figure 2, where we plot rows three through five of Table 4. The figure shows that in the earlier years both indicators under-predict the severely limited population, with the Combination Indicator being slightly closer. Starting in 1995, both indicators are indistinguishable in terms of the population they classify as severely limited, but they both now over-predict the population reporting a severe limitation by about three to

eight percent. Note that both indicators follow the trends of the true measure well – a feature not guaranteed by our approach. Since we only have eleven data points and are missing years in-between, regression analysis is not used to test this point more thoroughly.



Notes: All values are based on the index $S_\phi = (AD - BC)/[(A+B)(A+C)(D+B)(D+C)]^{1/2}$, where *A, B, C,* and *D* reflect the weighted numbers in the goodness-of-fit matrix, see text for details. The values for the "Logit Indicator" are derived from the Logit Estimation, for the "Combination indicator" we use a degree of disability of at least 53 or a satisfaction with health of at most 2 to define the disabled, see text for details. Missing years are 1986 (degree of disability not asked) and 1988 – 1992,1993 and 1994 (question about limitation not asked).

Source: Authors' calculations using SOEP 1984 – 2002.

Figure 1: Values of the Similarity Index S_ϕ for the Logit Indicator I^L and Combination Indicator I^C

5. Sample Comparisons

While Figure 2 and Table 4 give some insight into the sample proportions classified as severely limited by our indicators, we now show how close the samples are with respect to other characteristics. Table 5 reports sample characteristics split by disability status based on three different measures: the "true" measure of a reported severe limitation (columns 1 and 4), the Logit Indicator I^L (columns 2 and 5), and the Combination Indicator I^C (columns 3 and 6). The sample in Table 5 is restricted to one randomly chosen observation per person to omit multiple observations of a person. Our main interest is to compare characteristics of those characterized as severely limited, and those who are not, but we are also interested in the differences between these

Table 4: **Comparing Combination Measure and Logit Measure in their Goodness of Fit**[a)]

	All Years	1984	1985	1987	1992	1995	1996	1997	1998	1999	2000	2001
Correctly specified by												
Logit Indicator[b)]	91.67	90.65	91.28	91.27	92.17	91.73	91.49	91.70	91.21	91.60	92.45	92.42
Combination Indicator	91.12	90.35	90.90	90.65	91.37	90.73	90.97	90.88	90.91	91.42	92.08	91.73
Severely Limited by												
Logit Indicator[b)]	11.05	12.06	10.52	10.27	11.02	11.56	11.41	11.10	11.49	11.51	10.48	10.12
Combination Indicator	11.26	12.74	11.56	11.23	11.38	11.25	11.36	11.07	11.51	11.59	10.43	10.30
Reported a Severe Limitation	11.03	14.21	12.84	11.94	11.55	10.17	10.63	10.28	10.84	10.77	9.80	9.84
Observations	159,068	12,119	10,928	10,401	13,160	13,087	12,987	12,768	14,073	13,653	24,073	21,819

Notes:

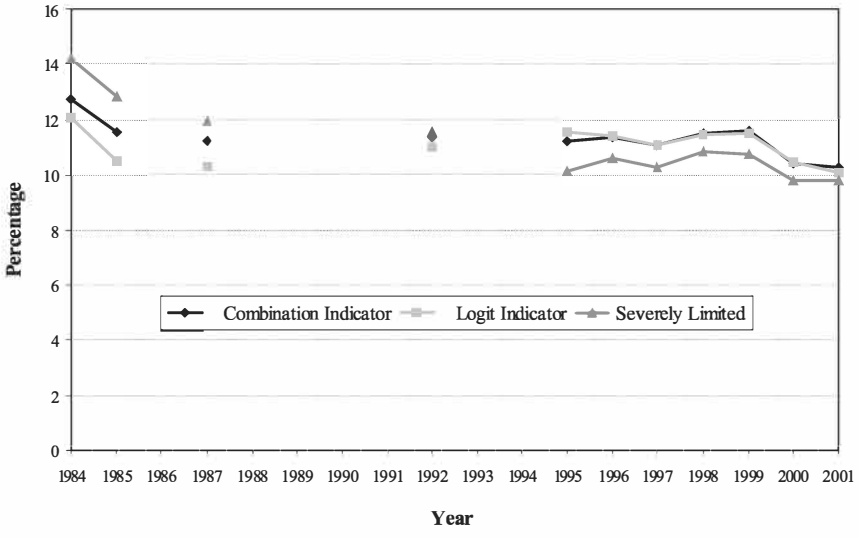
[a)] All values except the observations are weighted sample percentages. Missing years are 1986 (degree of disability not asked) and 1988 – 1992, 1993 and 1994 (question about limitation not asked).

[b)] The Logit Indicator is derived from the Logit Estimation.

[c)] A person is disabled by the "Combination Indicator" if reporting a degree of disability of at least 53 or a satisfaction with health of at most 2, see text for details.

Source: Authors' calculations using SOEP 1984 – 2002.

groups, not only to see in general what differences exist between them, but whether our indicators capture those differences.[14]



Notes: All values are weighted sample percentages. The "Logit Indicator" is derived from the Logit Estimation. A person is disabled by the "Combination Indicator" if reporting a degree of disability of at least 53 or a satisfaction with health of at most 2, see text for details. Missing years are 1986 (degree of disability not asked) and 1988 – 1992, 1993 and 1994 (question about limitation not asked).

Source: Authors' calculations using SOEP 1984–2002.

Figure 2: Comparing the Prevalence of Limitation by Indicator and Reported Limitation Status

Comparing health (row 1) across disability status (columns 1 – 3 vs. 4 – 6), we see that all measures do well in dividing the sample into healthy and unhealthy individuals. Within each measure, the severely limited have a much lower level of health satisfaction, a much higher degree of reported disability (row 2), are roughly 20 years older (row 3), and spent approximately seven to eight more nights at a hospital during the last year (row 4). A look at the reported severe (row 5) and slight limitations (row 6) serve as a consistency check. Both our indicators I^L and I^C identify a sample, where 93 and 92 percent report any limitation, respectively.

[14] While our restriction yields correct standard errors, the samples are still not independent and thus comparisons based on the standard errors would be misleading. However, as long as the correlation between the samples is positive, our estimates of the standard error of the difference between any two variables will be an upper bound. Since most of the differences are significant even without taking the correlations into account, they are likely to be true differences.

The fraction of females (row 7) is slightly larger in the severely limited samples, which is not surprising considering the higher average age in combination with the higher life expectancy for women. Individuals in the severely limited samples are less likely to be employed (row 8) at the time of the interview and, if employed, work fewer hours (row 9) and earn less (row 10). In addition, this group has a larger fraction of people without a high school degree (row 11), and fewer with a completed high school (row 12) or college (row 13) education, indicating a lower level of education in the severely limited group in general.[15]

A closer look at the differences in means between those who report being severely limited and those we estimate to be severely limited by our measures reveals that the "true sample" is more satisfied with their health (3.17) and registers a lower degree of disability (35.77) than our indicator samples. While the differences between age and nights spent in the hospital are small, there are differences of four percentage points in the fractions of females. This might be because fewer women participate in the workforce and hence are less likely to apply for official disability registration. Indeed, of employed people aged 21 – 58, only 42 percent are female. Among those who additionally register a disability, the fraction of females is only 36 percent. Since our indicators are based in part on the registration of a disability, fewer women are present in the severely limited subgroup specified by our indicators.

The differences we observe in employment probability are likely due to a number of factors: the older sample created by the true measure, the greater fraction of females and the slightly worse education. (The differences remain when restricting the samples to working-age individuals, although at an overall higher level of employment.) In addition, a finding by Burkhauser et al. (2002) may explain this difference as well: even though individuals have serious impairments, they are less likely to report a work limitation when employed. While health satisfaction could be affected similarly, it seems unlikely that the registration of a disability is and as a result, our indicators are less affected by this problem.

The difference in income through labor earnings could be explained by the employment benefits that individuals receive when they are registered as disabled through the Pension Office. Ceteris paribus, these benefits lead to a higher reservation wage, which results in higher compensation if employed. When the sample registered with a disability is split into those with a degree of disability of 30 percent and higher vs. those with a degree of 29 percent and below, we confirm this: the average income for those with a higher degree of disability is about five percent larger. The reliance on the registration measure then leads to higher incomes for the disabled samples, compared to the true measure.

---

[15] Note that the German level of education is translated into the American system: Realschule, Hauptschule, or no degree is labeled no high school, high school refers to Abitur, and college or higher is more than high school.

Table 5

Comparison of Samples Created by the Different Indicators[a)]

	Severely Limited			Not Severely Limited		
	Reported	Logit Indicator[b)]	Combination Indicator[c)]	Reported	Logit Indicator[b)]	Combination Indicator[c)]
Satisfaction with Health	3.171 (2.368)	2.253 (1.755)	2.754 (2.449)	7.152 (2.058)	7.237 (1.913)	7.188 (1.954)
Degree of Disability	35.77 (39.36)	45.76 (38.78)	49.21 (39.37)	3.43 (14.22)	2.43 (11.49)	1.88 (8.81)
Age	64.34 (16.34)	61.91 (16.67)	62.48 (16.71)	43.57 (18.21)	44.01 (18.55)	43.89 (18.46)
Nights Spent at the Hospital	9.50 (22.27)	9.28 (22.52)	9.02 (22.21)	1.36 (6.39)	1.44 (6.59)	1.46 (6.65)
Severely Limited	1.000 (0.000)	0.660 (0.474)	0.634 (0.482)	0.000 (0.000)	0.048 (0.215)	0.050 (0.218)
Somewhat Limited	0.000 (0.000)	0.274 (0.446)	0.284 (0.451)	0.285 (0.452)	0.250 (0.433)	0.249 (0.432)
Female	0.570 (0.495)	0.530 (0.499)	0.534 (0.499)	0.514 (0.500)	0.519 (0.500)	0.518 (0.500)
Currently Employed	0.175 (0.380)	0.218 (0.413)	0.205 (0.404)	0.576 (0.494)	0.568 (0.495)	0.570 (0.495)
Hours of Work Last Year[d)]	1945 (759)	2011 (808)	1956 (833)	2031 (708)	2028 (705)	2031 (704)
Labor Income[d)]	35.58 (37.81)	37.65 (29.09)	37.02 (28.55)	40.60 (34.87)	40.52 (35.29)	40.55 (35.30)
No High School	0.380 (0.485)	0.353 (0.478)	0.351 (0.477)	0.262 (0.440)	0.266 (0.442)	0.266 (0.442)
High School	0.533 (0.499)	0.544 (0.498)	0.534 (0.499)	0.576 (0.494)	0.574 (0.494)	0.576 (0.494)
More than High School	0.087 (0.282)	0.103 (0.304)	0.115 (0.319)	0.162 (0.369)	0.160 (0.366)	0.158 (0.365)
Observations	3565	3424	3443	33705	33846	33827

Notes:

a) All means are weighted by person-specific weights. Standard deviations in parentheses. Restricted to one randomly chosen observation per person. See text for details on the "Logit" and the "Combination" Indicators.

b) The Logit Indicator is derived from the Logit Estimation.

c) A person is disabled by the "Combination Indicator" if reporting a degree of disability of at least 53 or a satisfaction with health of at most 2, see text for details.

d) Hours Worked and Income refer to employed individuals only. Income in thousands of 2001 Deutsche Mark.

Source: Authors' calculations using SOEP 1984 – 2002.

A possible factor causing the differences in education levels could be that better educated people can more successfully get through the complicated administrative process of registering a disability with the Pension Office. One

would expect better educated individuals to be less deterred by administrative barriers, hence these individuals are more likely to be registered, and again, as our indicators are using the registered individuals, the samples we classify as disabled are better educated than those reporting a limitation.

The differences among the non-severely limited samples are relatively small – this is because the sample of those correctly specified as not having a severe limitation is large compared to those incorrectly specified. As such, the differences will not play a large role in any analysis using these samples.

Finally, we turn to the comparison of the samples defined by the indicators only. While the samples of not severely limited do not experience notable differences, within the severely limited sample, the performances of the indicators I^L and I^C slightly differ. While the sample identified as severely limited by I^L is closer to the true severely limited sample in the degree of registered disability and in nights at the hospital, the means of I^C and the true measure are closer for health satisfaction, age, and hours worked. For the other variables the means of the indicators are relatively close to one another.

Overall, the comparisons in Table 5 suggest that the benchmark measure of a severe limitation as well as our indicators are defining samples of individuals that experience health problems. The main differences are likely due to the registration of a disability which we argue could cause differences in gender, income, and education. We test the impact of these differences in a regression setting in the next section.

6. Regression Comparisons

We now show three regressions where we use each indicator to measure its influence on specific labor market outcomes and then compare the influences of the reported limitation with the one we predicted. In the first regression, we estimate the influence of a disability on yearly labor earnings (in logarithms), the second predicts its influence on current labor force participation. Finally we estimate a Heckman selection model of yearly labor earnings (in logarithms).[16]

The first two rows of Tables 6a, 6b and 6c show the estimated coefficients on reported and predicted severe limitation and their standard errors from the three sets of regressions. To test the coefficients from the predicted measures

[16] As is the case with the CPS and the PSID, the work limitation question in the SOEP is asked in real time but in the CNEF version of the SOEP all income questions including labor earnings are linked to income in the previous year. Since we are performing these regressions to more formally show how our indicators compare to the true measure, we will not discuss the magnitudes of the coefficients or their implications. For all regressions, we restrict the sample to working-age individuals (21 – 58) and allow for clustered error terms within each person's observations.

against the coefficient from the reported severe limitation, we provide a "strong" and a "weak" test. The "strong" test, which will reject the null hypothesis of equal values more often, assumes that the coefficient on the reported limitation is the true value (i.e. estimated without prediction error) and compares the coefficients of the predicted measures against this value. The "weak" test, less likely to reject the null hypothesis, takes both prediction errors into account, abstracts from the fact that the samples are identical and the indicators positively correlated, and provides a test of the difference of the estimated coefficients. The true measure for a significant difference will be somewhere between these two extremes.

Table 6a shows the outcomes from a regression of yearly labor earnings on disability status, where the sample is restricted to those people with positive labor earnings. (In addition, this regression controls for age, its square, hours worked last year, its square, education, gender and whether the person lived in the eastern part of Germany.) The reported severe limitation has a larger effect on labor earnings than the predicted indicators, with the coefficient of the combination measure I^C being closer to the true measure than to the coefficient obtained by I^L. This is confirmed in both the weak and the strong test: while neither of the differences is statistically significant at the 10 percent level, the null hypothesis that the difference of reported and predicted indicators is zero is less likely to be rejected for the combination measure. In addition, the predictive power of the combination indicator as measured by the R^2 is slightly better than that of the logit indicator.

Table 6a

Effect of Disability on Yearly Labor Earnings

	Severe Limitation		
	Is Reported	**By Logit Indicator**	**By Combination Indicator**
Coefficient	–0.0987 (0.0210)	–0.0682 (0.0220)	–0.0874 (0.0225)
R^2	0.5068	0.5066	0.5068
Strong Test (F)		1.9303	0.2530
Probability		0.1647	0.6150
Weak Test (t)		–1.0028	–0.3672
Probability		0.2413	0.3729
Observations	90219	90219	90219

Notes: Dependent variable is the logarithm of last year's labor income. The coefficient on each different disability indicator is shown. All regressions control for age, its square, hours and squared hours worked last year, education and gender. Standard errors (in parentheses) are clustered on the individual level. See text for the explanation on strong and weak tests.

Source: Authors' calculations using SOEP 1984–2002.

Table 6b suggests a similar pattern for the effect of a severe limitation on labor force participation, where now all working-age individuals are included in the logit estimation. (The estimation controls for age, its square, education, gender and whether the person lived in the eastern part of Germany.) Again, the reported severe limitation has a larger effect on labor force participation than the predicted indicators, and only the strong test for the Logit indicator I^L is rejected at a ten percent level of significance. The coefficient provided by I^C is slightly closer to the true measure than the logit indicator, which is also shown by the pseudo R^2.

Table 6b

Effect of Disability on Labor Force Participation

	Severe Limitation		
	Is Reported	**By Logit Indicator**	**By Combination Indicator**
Coefficient	–1.262 (0.067)	–1.147 (0.065)	–1.183 (0.067)
Pseudo R^2	0.0998	0.0983	0.0988
Strong Test (Chi^2)		3.0896	1.4113
Probability		0.0788	0.2348
Weak Test (Chi^2)		1.4994	0.7030
Probability		0.2208	0.4018
Observations	110707	110707	110707

Notes: Dependent variable is current labor market participation. The coefficient on each different disability indicator is shown. All regressions control for age, its square, education and gender. Standard errors (in parentheses) are clustered on the individual level. See text for the explanation on strong and weak tests.

Source: Authors' calculations using SOEP 1984 – 2002.

The previous results transfer into the estimates of a Heckman selection model in Table 6c. (The labor earnings equation includes age, its square, hours worked last year, its square, education, gender and whether the person lived in the eastern part of Germany; the selection equation includes gender, age and marital status.) Throughout, none of the null hypotheses can be rejected, but again we find that the combination indicator performs closer to the true measure in all instances.

Although our analysis is not (and can not be) exhaustive, the regressions suggest that our predicted measures do fairly well in comparison to the reported measure of severe limitations. All coefficients are smaller in magnitude than the true measure, where the combination measure performs slightly better than the logit measure in all regressions. However, this difference between the indicators is rather small and not statistically significant (using the strong test above).

Table 6c

Effect of Disability in a Sample Selection Model of Yearly Labor Earnings

	Severe Limitation		
	Is Reported	**By Logit Indicator**	**By Combination Indicator**
Income Effect	0.0967 (0.0256)	0.1203 (0.0251)	0.1052 (0.0260)
Selection Effect	–0.7286 (0.0431)	–0.7079 (0.0387)	–0.7227 (0.0407)
Log Likelihood	–5.49E+08	-5.49E+08	–5.49E+08
Tests: Income Effect			
Strong Test (Chi^2)		0.8777	0.1048
Probability		0.3488	0.7462
Weak Test (Chi^2)		0.4333	0.0553
Probability		0.5104	0.8140
Tests: Selection Effect			
Strong Test (Chi^2)		0.2871	0.0210
Probability		0.5921	0.8848
Weak Test (Chi^2)		0.1277	0.0099
Probability		0.7208	0.9207
Observations	110707	110707	110707

Notes: Dependent variable is logarithm of last year's labor market income. The coefficient on each different disability indicator is shown. The income equation also includes age, its square, hours worked last year (linear and squared), education and gender. The selection equation includes gender, marital status, and age. Standard errors (in parentheses) are clustered on the individual level. See text for the explanation on strong and weak tests.

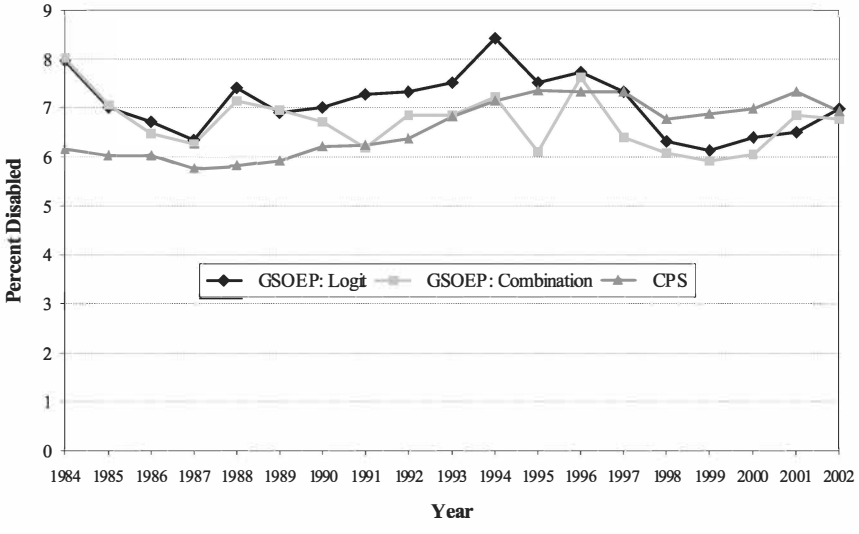
Source: Authors' calculations using SOEP 1984 – 2002.

Taken together, the analyses performed in section 4, 5, and 6 suggest that both indicators perform reasonably close to the true measure of severe limitation and are almost indistinguishable from one another. While the Logit measure performs slightly better in the initial evaluation of association, the combination approach is more successful in replicating the regression results based on the true measure. As such, we cannot choose one indicator over the other based on performance.

7. A First Application

In Figure 3 we use our consistently created measures of disability from the SOEP to estimate the prevalence of disability for working-age men (aged 21 – 58) in the western states of Germany between 1984 and 2002. We then com-

pare these levels and trends for the same working-age population of men in the United States using data from the CPS. As can be seen in Figure 3, the prevalence of disability among working-age men in Germany has with some variance been relatively stable over time. The mean is about 7 percent, and all years are within 20 percent of this value, most within 10 percent. Both measures have an outlier in otherwise relatively smooth series: the logit measure spikes in 1994, while the combination measure dips in 1995. Except for the early 1990s, both indicators are relatively close to one another. Comparing 1984 to 2002 we see a decline in the order of about one percentage point.[17]



Notes: Weighted means for the working-age population (21 – 58) in the western states of Germany and the United States. See Appendix Table 1 for exact numbers and sample sizes.

Source: Authors' calculations using SOEP and CPS 1984 – 2002.

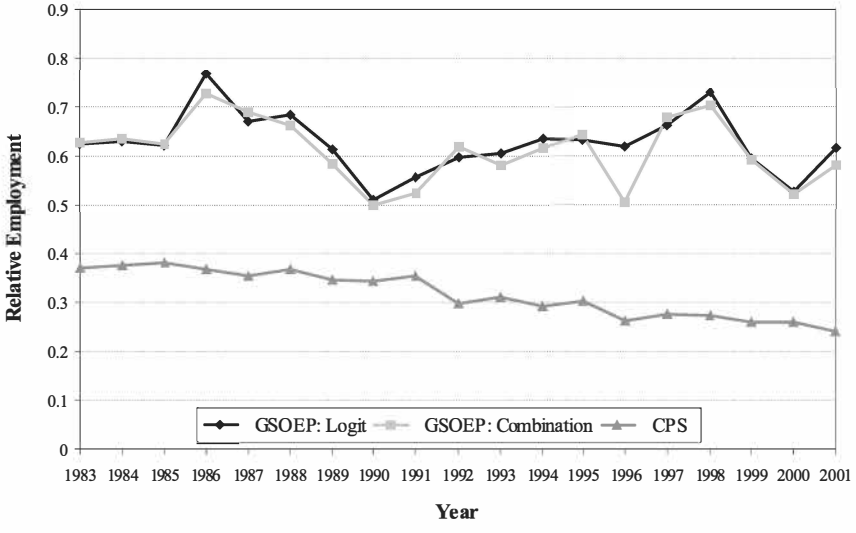
Figure 3: Prevalence of Disability among Working-Age Men (Aged 21 – 58) in Germany and the United States

In contrast the prevalence of disabilities among working-age men in the United States was relatively constant between 1984 and 1992, but then rose between 1992 and 1997 and drifted downward slightly thereafter. So over the entire period the prevalence of disability has risen from 6.17 percent in 1984 to 6.94 percent in 2002. One of the reasons why the variance in prevalence

[17] Parts of this effect might be an artifact of panel attrition. Especially in the first years in the panel, it seems likely that those individuals with health problems are more likely to leave the panel. This type of attrition is not accounted for in the weights.

rates across years in the United States is so much less over this period relative to that of Germany is that the sample size in the CPS ranged from 30,000 to 50,000 over this period, nearly ten times that of the SOEP (see Appendix Table 1).

In Figure 4 we measure the employment of working-age men with disabilities relative to that of working-age men without disabilities in both Germany and the United States. For Germany and the United States our employment indicator is total yearly hours worked (those with zero hours of work are included in this sample), a variable generated in the CNEF distribution of the SOEP (see Lillard / Wronski / Grabka, 2005). We use relative hours worked to control for general shocks to employment in both countries.



Notes: Weighted means for the working-age population (21 – 58) in the western states of Germany and the United States. See Appendix Table 2 for exact numbers and sample sizes.

Source: Authors' calculations using SOEP and CPS 1984 – 2002.

Figure 4: Employment (Hours Worked per Year) of Men with Disabilities Relative to Men Without Disabilities in Germany and the United States (Aged 21 – 58)

The first important point to note in Figure 4 is that in Germany, working-age men with disabilities work far more relative to their counterparts without disabilities than is the case in the United States. Over the entire period from 1983 through 2001, the relative hours worked of working-age men in Germany ranged from 50 to 76 percent while in the United States it ranged from 24 to 38 percent.

This finding of higher relative hours worked in Germany than in the United States for working-age men with disabilities is consistent with those of Burkhauser and Daly (1998) using the self-reported work limitation measure of disability in the SOEP data for the years it was available. It is also consistent with past cross national studies of disability policy based on aggregate administration data that find that the prevalence of disability transfer benefits among those aged 15 to 59 in Germany is far lower than in other OECD countries (see Aarts / Burkhauser / DeJong, 1996, 4, Table 1.1; Marin / Prinz / Queisser, 2004, 27, Figure 5).

It is highly unlikely that health differences are driving this consistently large difference in relative employment rates in the two countries found in Figure 4. It is much more probable that the social environment is different in the two countries. German disability policy has historically been much more work oriented than is the case in other OECD countries including the United States, with a much greater share of its public expenditures on disability going to vocational rehabilitation and other work based activities (see Aarts / Burkhauser / DeJong, 1996; Marin / Prinz / Queisser, 2004).

The second important point to note in Figure 4 is that the employment of working-age men with disabilities relative to their counterparts without disabilities follows different trends between 1983 and 2001 in both countries. There is a dramatic decline in relative employment of working-age German men between 1986 and 1990, the years just prior to German reunification, but this is followed by a steady increase in their relative employment through 1998. So between 1983 and 2001 there is on net a decline of only one percentage point.

In contrast, in the United States there is little change in the relative employment rates of working-age men with disabilities in the 1980s followed by a sharp decline between 1991 and 1992 and a general drift downward thereafter, yielding a thirteen percentage point differential between 1983 and 2001. While there is much greater variance in the yearly employment values in Germany than in the United States this is again most likely caused by differences in sample sizes rather than underlying market forces in the two countries.

It is likely that differences in the timing of changes in the social environment rather than in the severity of impairments in the two countries are at the root of these very differently timed downward trends in relative employment rates. In the United States, Acemoglu and Angrist (2001) argue that the Americans with Disabilities Act of 1990, first implemented in 1992, is responsible for this decline, while Autor / Duggan (2003) and Bound / Waidmann (2002) argue that it was caused by changes in the relative rewards offered by the Social Security Disability Insurance system (see Stapleton / Burkhauser (2003) for a more detailed discussion of this literature). Whatever the cause of this

decline in the United States, our cross-national comparisons show that the decline was limited to the United States. The relative employment of working-age German men with disabilities was rising over the 1990s.

8. Summary

This paper creates two yearly measures of disability for the SOEP that are comparable to measures of work limitation-based disability in United States datasets. We first identify a work limitation question that is similar to the one used in the CPS and PSID, which is available for some but not all years in the SOEP. We then use two health-related questions in the SOEP that are available for all years to create two alternative indicator variables that consistently measure work limitations. One indicator is based on a logit estimation using these two health-related questions, the other on a combination of these questions. We find the best definitions of each procedure using a similarity index, and then evaluate their performance against the original question by three trial regressions and a comparison of the different samples that are generated.

With these consistently measured disability variables we then estimate levels and trends in the prevalence rates of disability of working-age men in Germany. We also compare the employment of the working-age men with disabilities relative to their counterparts without disabilities in Germany and in the United States. We find that between 1983 – 2001 the relative employment of working-age men with disabilities in Germany is consistently higher than is the case for working-age men with disabilities in the United States. We also find that the substantial decline in the relative employment of working-age men with disabilities that occurred in the United States over the 1990s did not occur in Germany. Since it is improbable that variations in health between the two countries are responsible for these differences in levels and trends over the past two decades, it is likely that differences in the social environment faced in the two countries by working-age men with disabilities are at their root.

The measures of disability created in this paper make it possible to investigate the impact of different disability policies and policy changes in Germany alone or in comparison with the United States. Cross-national comparisons can be done by comparing yearly levels and trends in the two countries as reported here or by using the longitudinal structures of the SOEP and the PSID to dynamically assess how the onset of a disability affects employment and economic well-being.

References

Aarts, L. / *Burkhauser,* R. V. / *Dejong,* Ph. (eds.) (1996): Curing the Dutch Disease: An International Perspective on Disability Policy Reform, Aldershot, England.

Acemoglu, D. / *Angrist,* J. (2001): Consequences of Employment Protection? The Case of the Americans with Disabilities Act, Journal of Political Economy 109, 915 – 957.

Andreski, P. / *Gouskova,* E. / *McGonagle,* K. / *Schoeni,* B. (2005): An Analysis of the Quality of the Health Data in the Panel Study of Income Dynamics: 1999 – 2003, Downloaded April 2006: http: // psidonline.isr.umich.edu / Guide / report_on_health_qsv2.pdf.

Autor, D. / *Duggan,* M. C. (2003): The Rise in the Disability Rolls and The Decline in Unemployment, Quarterly Journal of Economics 118 (1), 157 – 206.

Banks, J. / *Kapteyn,* A. / *Smith,* J. P. / *van Soest,* A. (2004): International Comparisons of Work Disability, RAND Working Paper WR-155, April.

Bound, J. / *Waidmann,* T. (2002): Accounting for Recent Declines in Employment Rates among the Working-aged Disabled, Journal of Human Resources 37 (2), 231 – 250.

Burkhauser, R. V. / *Butrica,* B. A. / *Daly,* M. C. / *Lillard,* D. R. (2001): The Cross-National Equivalent File: A Product of Cross-National Research, in: I. Becker / N. Ott / G. Rolf (eds.) Soziale Sicherung in einer Dynamischen Gesellschaft Frankfurt am Main, 356 – 376.

Burkhauser, R. V. / *Daly,* M. C. / *Houtenville,* A. J. / *Nargis,* N. (2002): Self Reported Work-Limitation Data: What They Can and Cannot Tell Us, Demography 39, 541 – 555.

Burkhauser, R. V. / *Daly,* M. C. (1998): Disability and Work: The Experiences of American and German Men, FRBSF Economic Review 2, 17 – 29.

Burkhauser, R. V. / *Houtenville,* A. J. / *Rovba,* L. (2006): Accounting for the Declining Fortunes of Working-Age People with Disabilities, Working Paper, Cornell University.

Burkhauser, R. V. / *Weathers,* R. / *Schroeder,* M. (2006): A Guide to Disability Statistics from the Panel Study of Income Dynamics. Rehabilitation Research and Training Center on Disability Demographics and Statistics, Cornell University, Ithaca (NY).

Hotchkiss, J. L. (2003): The Labor Market Experience of Workers with Disabilities. The ADA and Beyond, Kalamazoo (MI). W.E, UpJohn Institute for Employment Research.

Kapteyn, A. / *Smith,* J. P. / *von Soest,* A. (2006): Dynamics of Work Disability and Pain, RAND Working Paper, WR-387, March.

Kolenikow, S. (2007): "Stata Modules and Tutorials from Stas Kolenikow." Accessed May 2007 at: http: // www.unc.edu / ~skolenik / stata.

Lillard, D. / *Wronski,* A. / *Grabka,* M. (2005), Codebook for the Cross-National Equivalent File 1980 – 2003, Accessed September 2006 at: http: // www.human.cornell.edu / che / PAM / Research / Centers-Programs / German-Panel / upload / GSOEP_CNEF_2.pdf.

Marin, B. / *Prinz,* Ch. / *Queisser,* M. (eds.) (2004): Transforming Disability Welfare Policies: Toward Work and Equal Opportunities, Aldershot, England.

Pearson, K. (1904): Mathematical Contributions To The Theory Of Evolution. XIII. On The Theory Of Contingency And Its Relation To Association And Normal Correlation, in: Drapers' Company Research Memoirs. Biometric Series, 1. Dulau and Co., London.

Pearson, K. / *Heron,* D. (1913): On Theories of Association, Biometrika 9 (1 / 2), 159 – 315.

Riphahn, R. T. (1999): Income and Employment Effects of Health Shocks. A Test Case for the German Welfare State, Journal of Population Economics 12, 363 – 389.

Sokal, R. R. / *Sneath,* P. H. A. (1973): Numerical taxonomy; the principles and practice of numerical classification, San Francisco.

Stapleton, D. C. / *Burkhauser,* R. V. (eds.) (2003): The Decline in the Employment of People with Disabilities: A Policy Puzzle, Kalamazoo, MI: W.E. UpJohn Institute for Employment Research.

Wagner, G. G. / *Burkhauser,* R. V. / *Behringer,* F (1993): The English Language Public Use File of the German Socio-Economic Panel, Journal of Human Resources 28 (2), 429 – 433.

Appendix

Appendix Table 1

Prevalence Rates of Disability for Working-Age Men in Germany and the United States

Year	Germany-SOEP			United States-CPS	
	% (Logit)	% (Combination)	*N*	% (Reported)	*N*
1984	7.95	8.01	4444	6.17	39152
1985	7.00	7.06	4022	6.02	39232
1986	6.73	6.47	3805	6.02	38742
1987	6.34	6.26	3866	5.77	38064
1988	7.40	7.15	3660	5.83	38345
1989	6.89	6.94	3559	5.93	35624
1990	7.01	6.71	3360	6.22	38558
1991	7.28	6.19	3469	6.25	38761
1992	7.31	6.86	3414	6.38	38473
1993	7.50	6.85	3237	6.83	38233
1994	8.41	7.21	3353	7.15	36788
1995	7.51	6.10	3530	7.36	36738
1996	7.72	7.61	3440	7.32	31881
1997	7.32	6.40	3311	7.33	32424
1998	6.31	6.08	3590	6.77	32496
1999	6.13	5.92	3384	6.88	32841
2000	6.40	6.06	6036	6.98	33552
2001	6.52	6.84	5356	7.33	32450
2002	6.98	6.76	5012	6.94	52926

Notes: All means are weighted by the respective individual weights and depict the fractions for working-age (aged 21 – 58) individuals. *N* refers to the sample size of non-missing observations with a positive weight.

Source: Authors' calculations using CPS 1984 – 2002 for the United States, SOEP 1984 – 2002 for Germany.

Appendix Table 2: **Mean Hours Worked per Year by those with and without Disabilities in Germany and the United States**

Year	Germany-SOEP								United States-CPS			
	Logit Measure				Combination Measure				Reported			
	Disabled	*N*	Not Disabled	*N*	Disabled	*N*	Not Disabled	*N*	Disabled	*N*	Not Disabled	*N*
1983	1259	354	2014	4090	1262	349	2014	4095	704	2504	1859	36648
1984	1229	277	1954	3745	1242	277	1954	3745	714	2563	1926	36669
1985	1217	237	1963	3568	1224	230	1960	3575	726	2545	1939	36197
1986	1490	245	1939	3621	1416	239	1944	3627	749	2534	1965	35530
1987	1292	248	1927	3412	1324	235	1923	3425	728	2448	1974	35897
1988	1332	229	1951	3330	1296	229	1954	3330	707	2262	1997	33362
1989	1175	220	1920	3140	1123	216	1921	3144	747	2489	2029	36069
1990	986	207	1928	3262	959	184	1918	3285	692	2542	2004	36219
1991	1075	216	1928	3198	1010	195	1929	3219	673	2597	1962	35876
1992	1123	210	1880	3027	1159	194	1872	3043	689	2731	1953	35502
1993	1086	237	1791	3116	1036	205	1786	3148	589	2645	1977	34143
1994	1114	208	1757	3322	1079	183	1750	3347	623	2596	2006	34142
1995	1113	221	1764	3219	1131	206	1762	3234	588	2199	2013	29682
1996	1115	192	1802	3119	916	167	1809	3144	610	2291	2020	30133
1997	1166	203	1762	3387	1193	195	1759	3395	535	2133	2038	30363
1998	1305	172	1790	3212	1260	157	1791	3227	570	2209	2058	30632
1999	1080	349	1819	5687	1077	304	1816	5732	565	2321	2060	31231
2000	963	307	1824	5049	953	288	1827	5068	534	2179	2056	30271
2001	1107	300	1798	4712	1047	278	1801	4734	521	3414	2010	49512

Notes: All means are weighted by the respective individual weights and depict the yearly hours for working-age (aged 21 – 58) men in the western states of Germany and the United States. *N* refers to the sample size of non-missing observations with a positive weight.

Source: Authors' calculations using CPS 1984 – 2002 for the United States, SOEP 1984 – 2002 for Germany.

Appendix 1. The German Disability Registration System

The German health system provides benefits to people who are officially registered as having a disability. A person can claim to have a disability when bodily functions, mental abilities, or the nervous system are expected to restrict a person's social life with a high probability for at least six consecutive months. Persons have to prove their disability status regularly with medical documents they send to the German Pension Office, which then determines the actual degree of disability on a scale from 1 to 100 percent. For example, a person with a heart condition that leads to a minor activity limitation (as evaluated by a doctor) is categorized as being disabled to a degree of 10 to 40 percent, moderate psychosis is classified between 50 to 70 percent, and a lost arm leads to a disability status of 100 percent. Combinations of impairments are possible, so that two minor impairments might lead to a severe classification.

All people who are officially registered and have a degree of disability of 20 percent or more get financial benefits, e.g., tax exemptions, deductions on insurance payments, rent allowances, etc. People categorized as disabled at a degree of 50 percent or higher are legally considered severely disabled, and are subject to special treatment in the labor market. For example, firms with more than 20 employees must employ at least 5 percent of severely disabled persons or otherwise pay a penalty of up to $280 per month for each quota place not filled with a disabled person. (This quota was 6 percent prior to 2000.) Individuals with a degree of disability between 30 and 50 percent can apply to be treated equally to a person with a degree of 50 percent or more, if they can prove that their disability affects their employment possibilities in the same way.

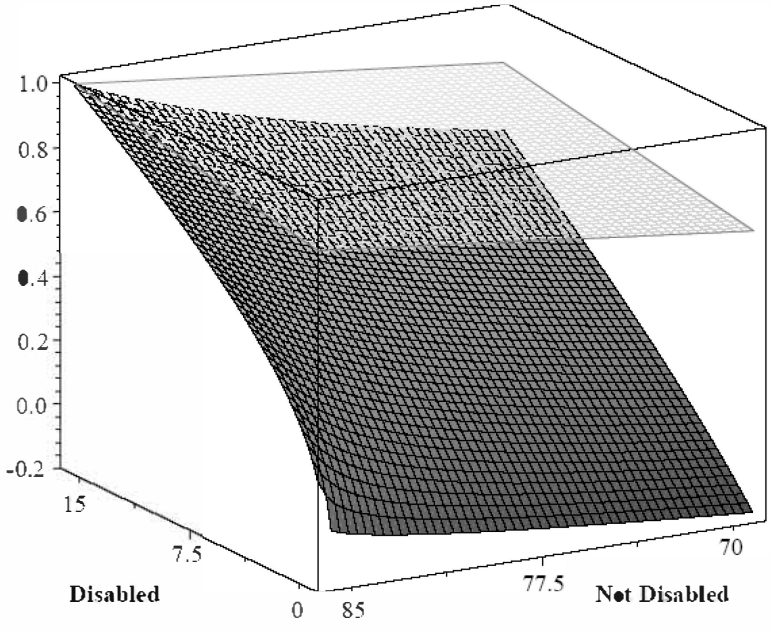
Appendix 2. Similarity Indices

The biology literature on classification of species uses similarity indices to evaluate how similar two species are (see for example Sokal and Sneath, 1973). This literature has developed numerous indicators to measure similarity based on the characteristics that are compared. As mentioned above, these indicators are based on the following goodness-of-fit matrix in some way:

		Characteristic A	
		No	**Yes**
Characteristic B	**No**	*A*	*B*
	Yes	*C*	*D*

There are instances, where only the match, i.e. cell *D*, would be of interest. An example taken from Sokal and Sneath is the classification of a species based on having wings – the fact that two species do not have wings does not add a lot to the similarity, and hence is not regarded as important. In our case, however, not having a disability is informative and hence the match between any definition and the true measure is important in both the disabled and the non-disabled population. In terms of the outcome matrix, we are interested both in *A* and *D*, and not in *D* alone. However, as mentioned in the text, we need to account for the differences in sample size in the disabled and the non-disabled groups, which the usual simple measure based on counting the matches is not able to achieve, whereas the coefficient S_ϕ, discussed in the text, achieves this goal.

To illustrate this point, we graphed both the simple index S_B and S_ϕ in a three-dimensional plot in Appendix Figure 1. We use the same hypothetical population as discussed in the text, which has 85 percent of people not disabled and 15 percent disabled. On the "Disabled" axis, the number of correctly specified non-disabled is held constant, and the number of correctly specified disabled is decreased. Along the "Not Disabled" axis, the number of correctly specified disabled is fixed, and the number of correctly specified non-disabled is reduced. Instead of graphing the full distribution of non-disabled, we restrict the maximum number of mis-classified non-disabled to 15 percent, i.e. cell A contains never less than 70 percent of the population. We let cell D vary between 0 and 15 percent, such that both axes have a range of 15 percent. The graph's upper left-hand corner has the highest amount of matches, i.e. no-one is mis-classified, whereas in the lower right-hand corner, overall only 70 percent of the population are specified correctly, i.e. no disabled person and specified as disabled and 15 percent of the not disabled are falsely specified as disabled.

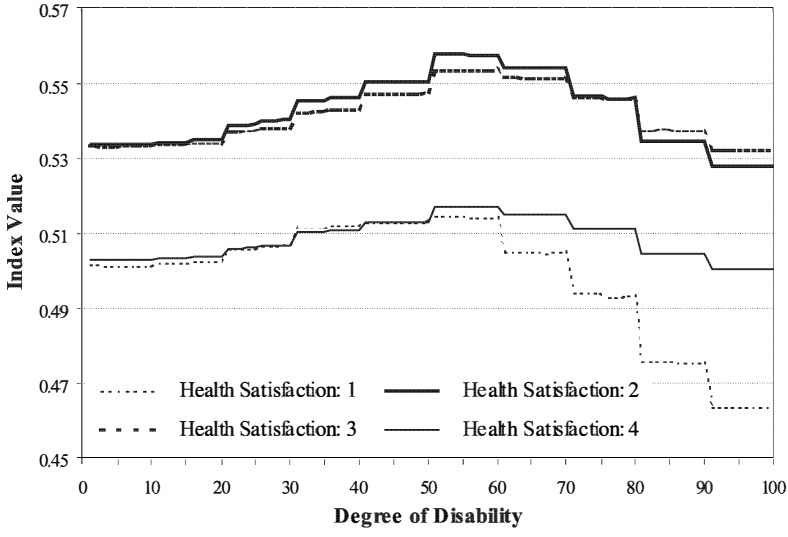


Notes: Authors' simulations based on the following formulas:
$S_B = (A + D)/(A + B + C + D)$
$S_\phi = (AD - BC)/[(A + B)(A + C)(D + B)(D + C)]^{1/2}$

Appendix Figure 1: Comparing the Basic Index S_B with Index S_ϕ

The two indices are depicted by the two shaded surfaces, where the surface of S_B is slightly lighter than the surface of S_ϕ. When comparing the two surfaces, we are not interested in the actual values the respective index produces, but rather in the differences in their slopes, since it is how mis-matches are valued against each other that is of interest in evaluating the indices.

The picture shows that the index S_ϕ is able to take into account the differences in sample size. This is best seen when comparing the marginal distributions starting from upper left hand corner. The slope in the coefficient is much steeper when decreasing the number of correctly specified disabled than when decreasing the number of correctly specified non-disabled. The basic index S_B has identical slopes for both axes. Thus, while index S_B would yield no difference in a five percent decrease in correctly specified disabled or non-disabled, using index S_ϕ would lead to choosing the one which has fewer disabled mis-classified.



Notes: Shown are the values for the S_ϕ index, where each line holds the level of health satisfaction constant and varies the degree of disability. As mentioned in the text, the health satisfaction levels are upper bounds, whereas the degree of disability is the lower bound. Any point in the graph represents the value of the index based on the weighted goodness of fit matrix between the severe limitation and the created indicator from the specific levels of health satisfaction and degree of disability, based on the whole sample. The index values shown are the ones with the highest values. Other levels of health satisfaction as well as the measures based on the joint presence of a certain level of health satisfaction and degree of disability are omitted in this graph.

Source: Authors' calculations using SOEP 1984 – 2002.

Appendix Figure 2: Values of S_ϕ by Level of Health Satisfaction and Degree of Disability