This is a repository copy of *Correction of misclassification error in disability rates*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/81264/

Version: Accepted Version

Article:

CORRECTION OF MISCLASSIFICATION ERROR IN DISABILITY RATES

AMANDA GOSLING and EIRINI-CHRISTINA SALONIKI*

School of Economics, University of Kent, Canterbury, Kent, UK

ABSTRACT

This paper examines misclassification error in survey estimates of disability. The results suggest that a significant number of those with a disability fail to be recorded as such in the British Household Panel Survey. In addition, the probability of a false positive is estimated as being very close to zero in all demographic groups. There is a strong bias in estimates of differences in rates of disability across groups but only a small effect on estimates of the difference in employment by disability status. Copyright © 2014 John Wiley & Sons, Ltd.

1. INTRODUCTION

This paper is about the misreporting of disability in survey data. We examine its extent, whether false positives are more likely than false negatives and the degree to which patterns in such errors may differ across sociodemographic groups. The impact on estimates of the prevalence of disability and its difference between men and women, the old and the young and between those with more or less education is considered. We also look at biases in differences in employment rates by disability status. The innovation of this paper is that misclassification error is identified without using supplementary information on the reliability of the disability measures (Bound, 1991; Kerkhofs and Lindeboom, 1995, 2009). Instead, it exploits multiple error-ridden survey measures of disability, as well as the relationship of employment to disability, and leverages assumptions that place restrictions on the correlation of errors across measures and employment status.

Misclassification errors can arise from problems in communication, because of the respondents' willingness to say what he or she believes the interviewer wants to hear and from the avoidance of answers that may be thought to arouse social disapproval. Respondents' answers may be influenced by their mood on the interview day or by their rapport with the interviewer (Lohr, 1999). More importantly, there are strong economic (Bound and Burkhauser, 1999) and psychological (Myers, 1982; Bowe, 1993; Hale, 2001) incentives to misreport disability status. Respondents may over report disability to take advantage of disability-related social benefits. Alternatively, they may not report being disabled in the fear of stigma and social exclusion.

The data used in this paper cover the last five waves (2004–2009) of the British Household Panel Survey (BHPS). There are quite a few disability-related variables in the BHPS such as choosing from a list of specific health problems that they have, and whether these problems limit daily activities or the amount of work a person can do (work-limiting vs non-work-limiting disability), a direct question on whether you consider yourself to be disabled or whether another member in the household considers you as disabled (and cares for you). Our procedure is predicated on the idea that differences in responses rates across these disability-related variables,

---

*Correspondence to: School of Economics, Keynes College, University of Kent, Canterbury, Kent, CT2 7RS, UK. E-mail: es343@kent.ac.uk, A.Gosling@kent.ac.uk

together with their differing impact on employment can, under certain assumptions, identify the extent and bias in misclassification errors. We require these variables to measure the same dimension of disability (i.e. that in the absence of misclassification error the responses would be the same) and that the misreporting rates for both variables are the same for the employed and the unemployed. The two variables in the BHPS that come closest to this are as follows. The first measure is constructed from a question on whether the respondent is being cared for by someone he or she lives with (CARE). The second measure comes directly from a question on whether the respondent considers themselves to be disabled or not (CONSIDER).

Our estimation procedure is non-parametric. We separate the data into cells defined by age, education, gender and year and estimate rates of false positives, false negatives, disability and so on separately for each cell.

We find that a significant number of those with a disability fail to be recorded as disabled. In addition, the probability of a false positive report of disability is estimated as being very close to zero in all our groups. There is also a strong bias in estimates of differences in rates of disability across groups. Interestingly, whereas there is a strong estimated bias in differences in employment rates by disability status for the second measure (CONSIDER), the bias using the first measure (CARE) is very small. We discuss the reasons for this in Section 4.

The structure of the paper is as follows. The next section presents our model and places it into the context of the existing literature on misclassification error together with a description of the estimation procedure. Section 3 describes the data, and Section 4 discusses the results. Section 5 concludes.

2. THE MISCLASSIFICATION PROBLEM AND ESTIMATION PROCEDURE

2.1. The misclassification problem

Consider a sample of individuals of which proportion d are disabled. These people are asked whether or not they are disabled. A proportion p of the disabled say they are not (false negative), and a further proportion q of the non-disabled say they are (false positive). Then, the proportion r saying they are disabled will be:

$$r = d(1 - p) + (1 - d)q \tag{1}$$

Note that $p = q$ is neither a necessary nor a sufficient condition for $r = d$. The latter is due to the fact that $r - d$ is a function of the number of people misreporting (dp and $q - qd$) and not just the misclassification probabilities. Also, the bias to d is a function of d itself. Smaller values of d are typically over stated whereas large ones under stated. This means that in contrast to what is often asserted, trends in d over time and differences in d across groups are also observed with error even if the underlying misclassification probabilities are constant.

The advantage for researchers is that knowledge of d (perhaps from another data source) can reveal something about the misclassification process. For example, Bross (1954) assumed that there are two methods of testing for disease prevalence. The first method is *unique* because it is available to the researcher and is without measurement error ('gold standard'). The second method is again available to the researcher but with measurement error. Then, using the concept of the fourfold table (2×2 table) together with the observed sample proportions, and assuming that the misclassification probabilities are equal and each less than 0.5, he showed that they can be calculated directly. On the basis of this, he also proposed a way to derive the bias between the 'gold standard' method and the second method with the measurement error. An extension of his proposed model came later on from Tenenbein (1972) who allowed the two methods to follow a multinomial distribution. In econometrics, Bollinger (1996) showed that the assumption $p = q$ could be relaxed and bounds on the unknown parameters could be obtained. A similar approach, proposed initially by Horowitz and Manski (1995) and extended by Kreider and Pepper (2007, 2008), imposes further distributional and functional assumptions and provides bounds for the estimated coefficients of the variables measured with error.[1]

---

[1]This approach is known as monotone instrument variable bound (MIV).Kreider and Pepper (2007) found that under relatively weak non-parametric assumptions, non-workers appear systematically to over report being disabled.

But what if there is no gold standard available to the researcher? Bollinger (1996) shows that bounds to p and q can give informative (i.e. strictly above zero or below one) bounds to d. Hui and Walter (1980) also show that if there are two methods of testing for disease prevalence, *none* of them needed to be a gold standard provided that the researcher is prepared to make assumptions about the joint misclassification probabilities in the two measures. To see this, let us write down the problem with two measures:

$$r_1 = d(1 - p_1) + (1 - d)q_1 \tag{2}$$

$$r_2 = d(1 - p_2) + (1 - d)q_2 \tag{3}$$

If the measures are available for the same sample, we also have

$$r_{both} = d(1 - p_1 - p_2 + p_{both}) + (1 - d)q_{both} \tag{4}$$

where r_{both} is the proportion of reporting being disabled in response to both measures, p_{both} is the probability a disabled person reports not being disabled on both measures and q_{both} is the probability a non-disabled person reports being disabled on both measures.

We can think of equations (2–4) as a nonlinear system of three simultaneous equations with five unknowns $\{p_1, q_1, p_2, q_2, d\}$. Restrictions can be then made to the system to obtain identification. Possible restrictions are as follows: Q2

$$p_1 = q_1 \text{ and } p_2 = q_2 \text{ (no bias)} \tag{A.1}$$

$$p_1 = p_2 \text{ and } q_1 = q_2 \text{ (symmetry)} \tag{A.2}$$

$$p_{both} = p_1 p_2 \text{ and } q_{both} = q_1 q_2 \text{ (no correlation)} \tag{A.3}$$

$$q_1 = q_2 = 0 \text{ (no over reporting)} \tag{A.4}$$

$$p < \frac{1}{2} \text{ and } q < \frac{1}{2} \tag{A.5}$$

For example, we could invoke (A.1) and (A.3) (as in Hui and Walter, 1980) together giving us an identified system of three equations with three unknowns $\{p_1, p_2, d\}$. However, because the system is nonlinear (quadratic) an assumption such as (A.5) may be necessary to pin down the parameters.[2] This system can be solved using nonlinear least squares or maximum likelihood.

It is important to mention that *not* all assumptions (A.1)–(A.5) are plausible in the context of the measurement of disability. The no bias would seem particularly strong. There is no good reason to expect the probability a disabled person reports not being disabled to equal the probability a non-disabled person reports being disabled. If stigma is important, then one would think that $p > q$, whereas if people over report disability to gain access to benefits, then one would expect $p < q$. Next, symmetry requires that the misclassification process is the same across both variables, not only is this not likely but it is also rejected by our data ($r_1 \neq r_2$).

Our extension allows identification with a weaker set of assumptions on p and q by using the joint distribution of the two discrete measures of disability and another outcome, which in our case is employment. Let upper case letters represent the discrete variable counterparts to the sample proportions indicated by lower case letters, for example, $D = 1$ indicates being disabled, $R_1 = 1$ indicates reporting disability by measure 1 and $E = 1$ indicates employment.

---

[2]This is the discrete equivalent of solving the measurement error problems with continuous variables using IV.

If employment depends on disability status such that $\Pr(E=1|D=1) \neq \Pr(E=1|D=0)$, then $\Pr(E=1|R_1, R_2)$ will depend on $\Pr(D=1|R_1, R_2)$. In addition, if the probability that a disabled person reports being disabled is independent of employment status, $\Pr(R_1=1|D=1, E=1) = \Pr(R_1=1|D=1)$, then the conditional probability of employment will be a known function of the ps, qs and d.

$$e_{R_1} = e_{ND} + (e_D - e_{ND}) \frac{d(1-p_1)}{d(1-p_1) + (1-d)q_1} \qquad (5)$$

where $e_{R_1} = \Pr(E=1|R_1=1)$, $e_{ND} = \Pr(E=1|D=0)$ and $e_D = \Pr(E=1|D=1)$.

As there are eight possible combinations of (E, R_1 and R_2), this gives seven equations and two more unknowns: e_{ND} and e_D. Without further restrictions, this system, similar to that of Hui and Walter, is under identified. Unlike theirs however, we only need to impose *one* restriction to achieve identification. We believe that the weakest assumption we can invoke is that of no correlation between the measures (refer to (A.3)). We then consider this assumption together with the others we made, in more detail.

The assumptions we made in order to identify the parameters of interest are as follows:

Answers to CARE and CONSIDER would be identical in the absence of measurement error; thus, they do not only pick up the same dimension but also pick up the same extent of disability.

As mentioned in Section 1, we have chosen the variables in the BHPS for which this assumption is most plausible. Although the assumption of dimension is probably uncontroversial, it is likely that those being cared for in the home are 'more' disabled than those who are not. In support of our approach, many disabled people do live alone or in sheltered accommodation. Thus, it is possible that not all those with the same condition will be cared for by a relative living with them.

The misclassification process in the two variables is independent.

Support for such an assumption comes from the fact that the variables are coded using questions asked of different individuals. It is however, possible that family or household differences in attitudes may make this assumption invalid.

The misclassification process is independent of employment. Thus, those working are not more likely to under or over report their status compared with those not working.

This is, we believe, our most problematic assumption. Indeed, much research (Bound, 1989, 1991; Kreider, 1999; Kreider and Pepper, 2007) has discussed the issue of 'justification bias', meaning that respondents may describe themselves as disabled in order to better justify why they are not working. It is for this reason that we have not used the questions in the BHPS related to limitations in daily activities or work limitations caused by the disability, that is, 'Does your health in any way limit your daily activities compared to most people of your age?' and 'Does your health limit the type of work or the amount of work you can do?'. Our results however do need to be interpreted cautiously.

Although our procedure does involve making (in some cases *heroic*) assumptions about the underlying process of misclassification, it is important to remember that these are weaker than those hitherto used (e.g. that of no bias). Future work by others may be able to exploit different variables where such assumption is more likely to be valid.

The previous analysis abstracts from other determinants of E and D and the probability of misreporting. These may be of interest for their own sake, or it might be that the identification restrictions only hold conditional on some set of observed variables. If these variables (Xs) are continuous, then we would need to incorporate assumptions on the joint distribution of X, E and D. Far easier is the case of discrete Xs; here, the researcher simply needs to estimate the unknowns within each cell defined by X.

2.2. Estimation procedure

As mentioned earlier, we exploit two subjective indicators of disability. For the first measure, a disabled person is identified from another member in the household who looks after him. In particular, each respondent is asked

the general question 'Is there anyone living with you who is sick, disabled or elderly whom you look after or give special help to (for example, a sick, disabled or elderly relative/husband/wife/friend, etc)?'. If he responds 'Yes' to this question, he is also asked 'Who is the person/people you look after?'. Then, we match the household identification number that they mention with the unique personal identification number, and we trace back each respondent's perceived disability status. For the second measure, the respondent is classified as disabled if he has responded positively to the question 'Do you consider yourself to be a disabled person?'.

Once the no correlation restriction is invoked, the system has seven unknown parameters $\{p_1, p_2, q_1, q_2, d, e_D, e_{ND}\}$. Each observed proportion (e.g. $\Pr(R_1 = 1|X)$) will be a known function of these unknown parameters, which could be estimated by maximum likelihood (as Hui and Walter (1980) did), nonlinear least squares or grid search.

The fact that all the unknowns are probabilities and thus bounded in the $\{0, 1\}$ interval makes the third method appropriate. However, Monte Carlo simulations on data similar to that used in our study reveal that the results are very similar to those estimated using nonlinear least squares with the *crucial* difference that results are obtained in a matter of seconds than in hours. We leave a comparison of nonlinear least squares with maximum likelihood methods to future work.

The nonlinear procedure is simple. It minimizes

$$\sum_{k=1}^{7} \left(O_k - \hat{O}_k \right)^2 \tag{6}$$

where O_k is the proportion of individuals in the cell observed with outcome k (e.g. reporting themselves to be disabled on the first measure).

There are many possible sets of outcomes relating to a different way of partitioning each group. The set used in our estimation is given in the succeeding text together with the formulae linking them to the unknown parameters.

$$O_1 = \Pr(R_1 = 1) : \hat{O}_1 = \hat{d}(1 - \hat{p}_1) + (1 - \hat{d})\hat{q}_1 \tag{7}$$

$$O_2 = \Pr(R_2 = 1) : \hat{O}_2 = \hat{d}(1 - \hat{p}_2) + (1 - \hat{d})\hat{q}_2 \tag{8}$$

$$O_3 = \Pr(R_1 = 1, R_2 = 1) : \hat{O}_3 = \hat{d}(1 - \hat{p}_1)(1 - \hat{p}_2) + (1 - \hat{d})\hat{q}_1\hat{q}_2 \tag{9}$$

$$O_4 = \Pr(E = 1|R_1 = 1) : \hat{O}_4 = \frac{\hat{e}_D\hat{d}(1 - \hat{p}_1) + \hat{e}_{ND}(1 - \hat{d})\hat{q}_1}{\hat{O}_1} \tag{10}$$

$$O_5 = \Pr(E = 1|R_2 = 1) : \hat{O}_5 = \frac{\hat{e}_D\hat{d}(1 - \hat{p}_2) + \hat{e}_{ND}(1 - \hat{d})\hat{q}_2}{\hat{O}_2} \tag{11}$$

$$O_6 = \Pr(E = 1|R_1 = 1, R_2 = 1) : \hat{O}_6 = \frac{\hat{e}_D\hat{d}(1 - \hat{p}_1)(1 - \hat{p}_2) + \hat{e}_{ND}(1 - \hat{d})\hat{q}_1\hat{q}_2}{\hat{O}_3} \tag{12}$$

$$O_7 = \Pr(E = 1|R_1 = 0, R_2 = 0) : \hat{O}_7 = \frac{\hat{e}_D\hat{d}\hat{p}_1\hat{p}_2 + \hat{e}_{ND}(1 - \hat{d})(1 - \hat{q}_1)(1 - \hat{q}_2)}{1 - \hat{O}_1 - \hat{O}_2 + \hat{O}_3} \tag{13}$$

Monte Carlo simulations (available upon request) have shown this system to be consistent so long as the sample is reasonably (>200) large. The effects of sample size are exacerbated by d. For example, if $d = 1\%$ and the sample size is 200, then there will be a high chance that there will be one or zero people in the sample with which to estimate the ps and e_D. We thus restrict some of the analysis to groups where the sample size is reasonably large and the underlying risk of disability is not too small.

To obtain p-values and confidence intervals, we bootstrapped the sample 350 times and conducted the estimation on each of these bootstrapped samples.

The above has abstracted from covariates, but we are interested in how the unknown parameters (es, ps, qs [Q4] and d) vary by age, gender and qualifications. We thus obtain separate Os for each age, gender and qualification group in the following way.

We define seven age groups (25, 30, 35, 40, 45, 50 and 55), four education groups (degree and above, basic school qualifications, higher school qualifications and no qualifications) and two gender groups (male and female), and centre our data round 1 year (2007). Overall, we get 56 cells. We calculate our observed outcomes (sample proportions and employment probabilities) using a non-parametric procedure with a Gaussian kernel; that is, for $O_1 = r_1$,

$$\Pr(R_1 = 1|x_j) = O_{1j} = \sum_{i=1}^{n} I(R_{1i} = 1)K_j(x_i) \frac{1}{\sum_{i=1}^{n} K_j(x_i)} \tag{14}$$

where x is a vector of observed individual characteristics and time variables (age, gender, qualifications and year), and $K_j(x_i)$ is a kernel function based on the 'distance' between x_i and x_j.

For example, if x_j is age $= 45$, sex $=$ male and qualifications $= 1$, then

$$K_j(x_i) = I(sex = male)I(qual = 1)\phi\left(\frac{age_i - 45}{3}\right)\phi\left(\frac{year_i - 2007}{2}\right).$$

3. DATA

The data used in this paper are from the BHPS covering five waves between 2004 and 2009 because these are the most consistent data spells available on health and employment rates. For the purpose of this analysis, we selected all those aged between 21 and 59 (inclusive) years who gave valid responses to the questions on disability (first and second measures), education, gender, age and employment status. This gave a total sample of 45,457 observations over the 6 years. A respondent is reported disabled on the first measure if another member of the household responded positively to the question 'Is there anyone living with you who is sick, handicapped or elderly whom you look after or give special help to (for example, a sick or handicapped (or elderly) relative/husband/wife/friend, etc.)?', followed by the question 'Who is the person/people you look after?'. A person is reported disabled on the second measure if he or she responded positively to the question 'Can I check, do you consider yourself to be a disabled person?'. Our estimation procedure is non-parametric. Table I shows the prevalence of reported measures [T1] of disability in this sample. The risk of disability is not dramatically different between men and women, but both measures show it to be falling with education and increasing with age. The first measure has a much smaller prevalence and most of those reported as disabled on this measure also report it on the second.

Table I. Prevalence of reported disability

	Respondent is cared for by other household member (measure 1)	Respondent reports being disabled (measure 2)	Measure 1 and 2	Observations
Whole sample (21–59 year old)	2.57%	6.75%	2.09%	$N = 45\,547$
All men	2.37%	6.69%	2.03%	$N = 20\,689$
All women	2.73%	6.79%	2.14%	$N = 24\,858$
With degrees and above	0.77%	2.79%	0.69%	$N = 9089$
With higher school qualifications	1.66%	4.88%	1.26%	$N = 13\,644$
With lower school qualifications	2.40%	5.84%	1.90%	$N = 14\,859$
With no qualifications	6.47%	16.14%	5.44%	$N = 7955$
Those with age < 50	1.98%	4.98%	1.57%	$N = 34\,729$
Those with age ≥ 50	3.93%	11.10%	3.27%	$N = 15\,579$

Under the assumption that the first and second measures pick up the same underlying construct of disability, this table provides strong evidence for misclassification error. To see this, note that if there is no misclassification error, then no individual will be reported as disabled on one measure and not on another. Under the null of no misclassification error, there is a zero probability of sampling anybody in the 'off-diagonal' cells (($R_1 = 0$, $R_2 = 1|D = 1$) or ($R_1 = 1, R_2 = 0|D = 1$)), and the fact that we do, indicates that misclassification error is present or else that the assumption that the two measures correspond to the same latent disability construct does not hold (see the discussion at the end of Section 2.1).

It should be noted that symmetry assumption is rejected as the number of people reported as disabled on the first measure is significantly different from the number of those reporting disability on the second.

4. RESULTS

4.1. Misclassification probabilities

It is easy to see that the overall extent of misclassification error can be gauged by looking at reporting inconsistencies. Such cross tabs can however, tell us nothing about the direction of the biases (and whether over reporting is more likely than under reporting). This can be deduced from changes in the probability of employment with each definition of disability status. To see this, consider the case where $q_1 = q_2 = 0$ and p_1 and p_2 are both large with $p_1 > p_2$. In this case, all those reporting themselves to be disabled will in fact be so, and the measured employment proportion of the disabled will be the same if we use measure 1, measure 2 or restrict it to those recorded as disabled under both measures. Conversely, as p_1 and p_2 are both large, the proportion of those in the recorded able-bodied group will include many workers that are in fact disabled. As $p_1 > p_2$, their employment rates will change when we move from those able bodied under measure 1 (including many disabled workers), to those able bodied under measure 2 (including fewer disabled workers) and finally to those able bodied under both measures as $p_1 p_2$ will always be smaller than both p_1 and p_2.[3]

As predicted earlier, our estimation procedure fails to find a solution for groups with particularly low rates of disability (young workers with high levels of education). We thus have to exclude those under 40 years old with higher school or degree and above qualifications from the results. Table II presents the misclassification probabilities. For any sociodemographic group, the probability of under reporting is greater than the probability of over reporting, supporting the idea that avoidance of stigma and social exclusion attached to disability are stronger motivations for not reporting disability than are the positive incentives for reporting disability provided by disability benefit entitlements and justification for not working ($p > q$). The false negatives are above 0.5 on average, but the false positives are by far very low and close to zero. When comparing the two measures, false negatives in the second measure are smaller than the ones in the first measure ($p_2 < p_1$)—they do not exceed 17% on average. Over reporting in the second measure is slightly greater than over reporting in the first for all groups, except women and those with higher school qualifications, however still very low ($q_1 < q_2$ and qs approximate zero). No clear inferences can be made for any differences in the misclassification probabilities between men and women. Finally, when using the second measure to infer disability, the higher school or unqualified people are more likely to be misclassified compared with those with degrees and above.

The proportion of false negatives is greater than the proportion of false positives in more than 99% of all bootstrap samples for both measures and all sociodemographic groups, with the exception of the highest educated for measure 2. The proportion of false negatives using care received from a household member (measure 1) is greater than the proportion of false negatives based on direct reporting of disability status (measure 2) for more than 99% of bootstrap samples. The proportion of false positives by measure 1 is less than the proportion of false positives by measure 2 for over 99% of the bootstrap samples.

T2

---

[3]The logic is shown more formally in the appendix.

Table II. Misclassification probabilities

| | False negative | | False positive | |
	p_1	p_2	q_1	q_2
Whole sample (21–59 year olds)	0.593	0.174	0.0014	0.015
All men	0.631	0.160	0.0017	0.018
All women	0.561	0.186	0.0010	0.013
With degrees and above	0.509	0.116	0.0010	0.016
With higher school qualifications	0.705	0.171	0.0010	0.010
With lower school qualifications	0.559	0.203	0.0009	0.015
With no qualifications	0.621	0.142	0.0028	0.019
Those with age ≥ 50	0.646	0.172	0.0014	0.019

Respondent is cared for by other household member (measure 1) and respondent reports being disabled (measure 2). Hence, p_1 is false negative by measure 1, q_2 is false positive by measure 2 and so on.

4.2. Corrected estimates of disability prevalence

Figures 1 and 2 show the reported and corrected estimates of disability rates of men and women by age and for the different qualification groups. The estimated rates correspond to $\hat{d}$ and are the reported rates purged of the estimated false positives and negatives across the two measures. It is important to remember that the bias in the reported rates depends on the relative *numbers* of false positives and negatives and solely on the size of the *p*s and the *q*s. Note that **F1 F2**
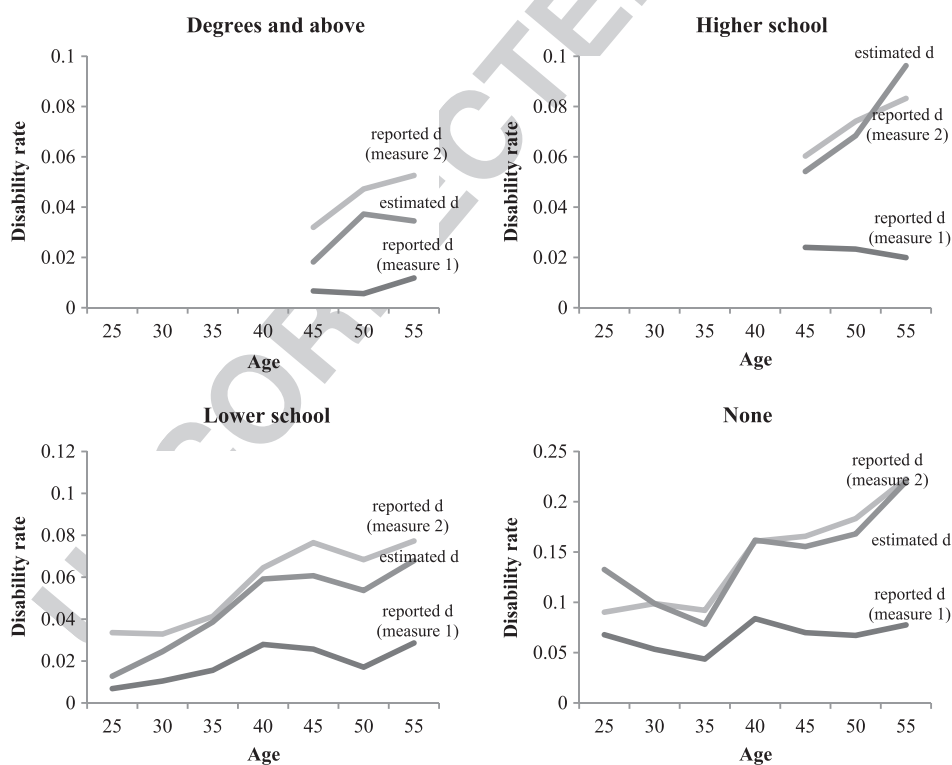
Note: Respondent is cared for by other household member (measure 1) and respondent reports being disabled (measure 2)

Figure 1. Reported and estimated disability rates of men
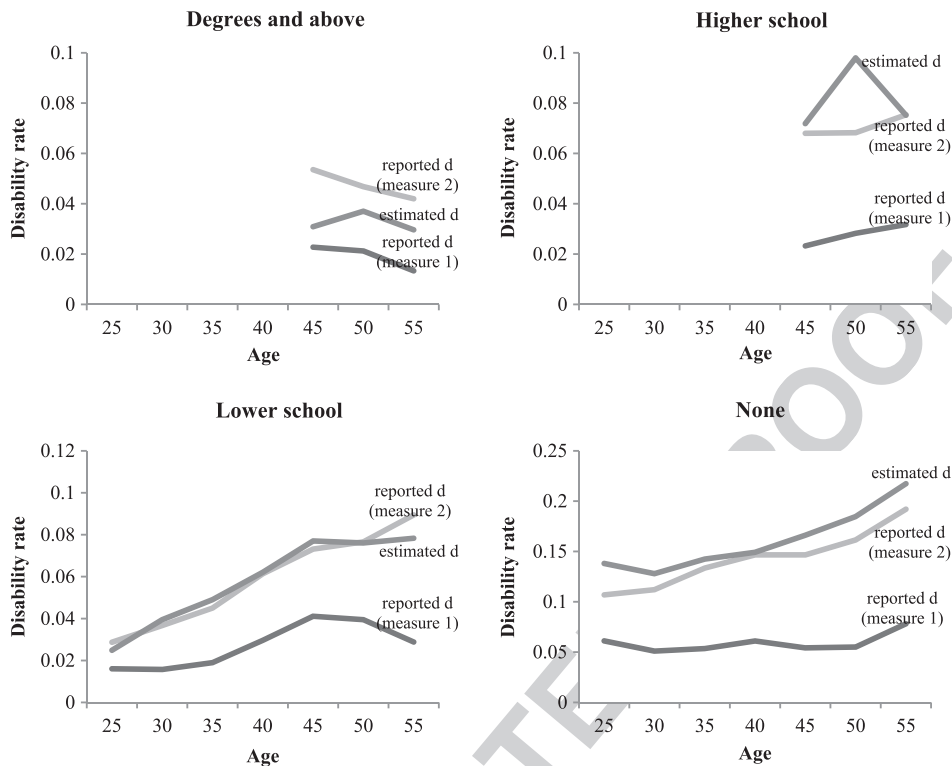
Note: Respondent is cared for by other household member (measure 1) and respondent reports being disabled (measure 2)

Figure 2. Reported and estimated disability rates of women

$$r - d = d(1 - p) + (1 - d)q - d = q - d(p + q) \tag{15}$$

For men younger than 55 years, the estimated disability rates are always smaller than those reported using measure 2 (CONSIDER), and larger than that using measure 1 (CARE). This is to be expected with qs very close to zero, d less than 0.5 and with higher false negatives in the first measure than the first (it is also true that $\mathbf{p_2 < d < p_1}$). In addition, for men of the same age group, the bias in measure 1 appears to be larger (in absolute terms) than that in measure 2. The reason for this can be seen from equation (15). The bias to the first measure is negative as q_1 approximates zero, but it depends on dp_1, which itself is small but much bigger than dp_2.

For women, the bias to the disability over the life cycle is more acute in the first measure. This again follows the intuition of equation (15) (tiny qs, but dp will be bigger under measure 1). The estimated disability rates of women vary more by age in the second measure. In particular, the estimated disability rate is above the reported ones for those with higher school or no qualifications, but the bias is greater for women at the age of 50 years with higher school qualifications. In the basic qualifications group, the estimated prevalence of disability for women at the age of 45 years lies above the reported rate in the second measure. This is because for this specific group, the estimated disability rate is greater than any of the ps. For respondents with degrees and above, the pattern for both genders is very similar.

Overall, it can be said that for both genders, the estimated disability rate is somewhere between the two reported rates, and it increases over the life cycle. The patterns are very similar for both genders. The estimated disabled do not account for more than a quarter of the total sample with a low of approximately 6% for men with higher qualifications and 3% for women with no qualifications. The biggest deviations of the estimated disability rate from the reported ones are observed mostly at the age of 55 years for those with higher or no qualifications at all.

4.3. Estimated employment probabilities

For ease of comparison before and after correcting for the misclassification probabilities, we define three differences (observed and estimated).
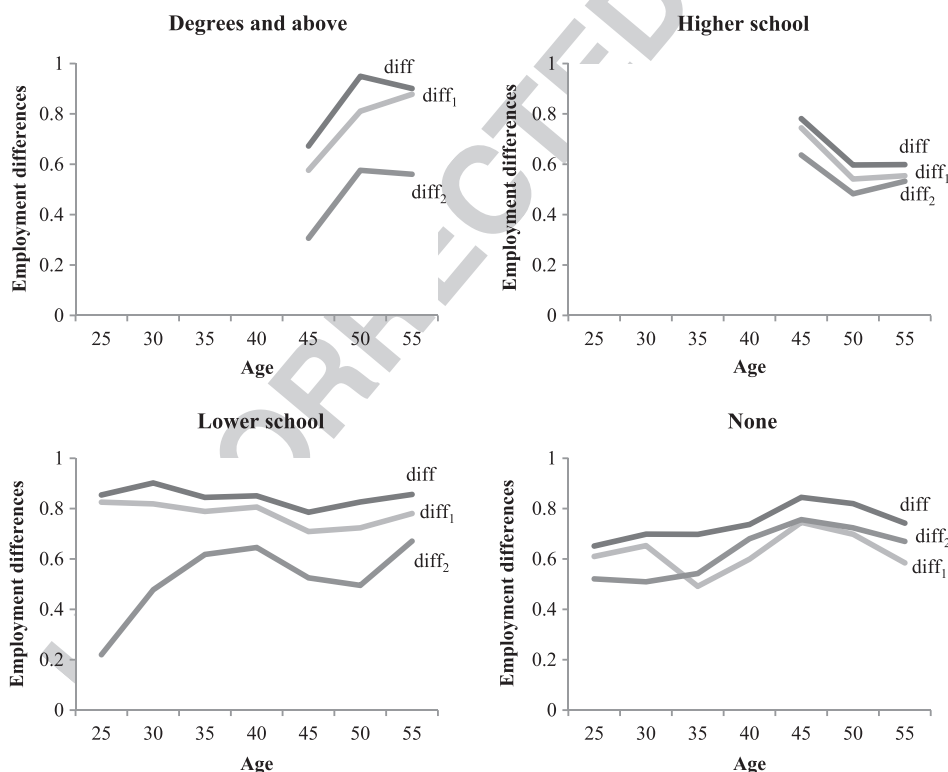
The estimated difference is

$$diff = e_{ND} - e_D = \Pr(E = 1|D = 0) - \Pr(E = 0|D = 1) \tag{16}$$

and the observed differences are

$$diff_1 = \Pr(E = 1|R_1 = 0) - \Pr(E = 1|R_1 = 1) \tag{17}$$

$$diff_2 = \Pr(E = 1|R_2 = 0) - \Pr(E = 1|R_2 = 1) \tag{18}$$

These differences are presented in Figures 3 and 4, separately for men and women, age and qualification **F3 F4** group. For those with degrees and above or higher school qualifications, we again restrict the sample to older adults older than 40 years old. The differences in employment rates by reported disability status are very large. For the least educated men reporting themselves disabled (measure 2), their rate of employment is 40–70 percentage points below that of their reported able-bodied counterparts.

Note: Estimated difference in employment rates by disability status (diff), reported difference in employment rates by disability status for measure 1 (diff₁), reported difference in employment rates by disability status for measure 2 (diff₂)

Figure 3. Difference in employment rates by disability status based on reported and estimated disability—men
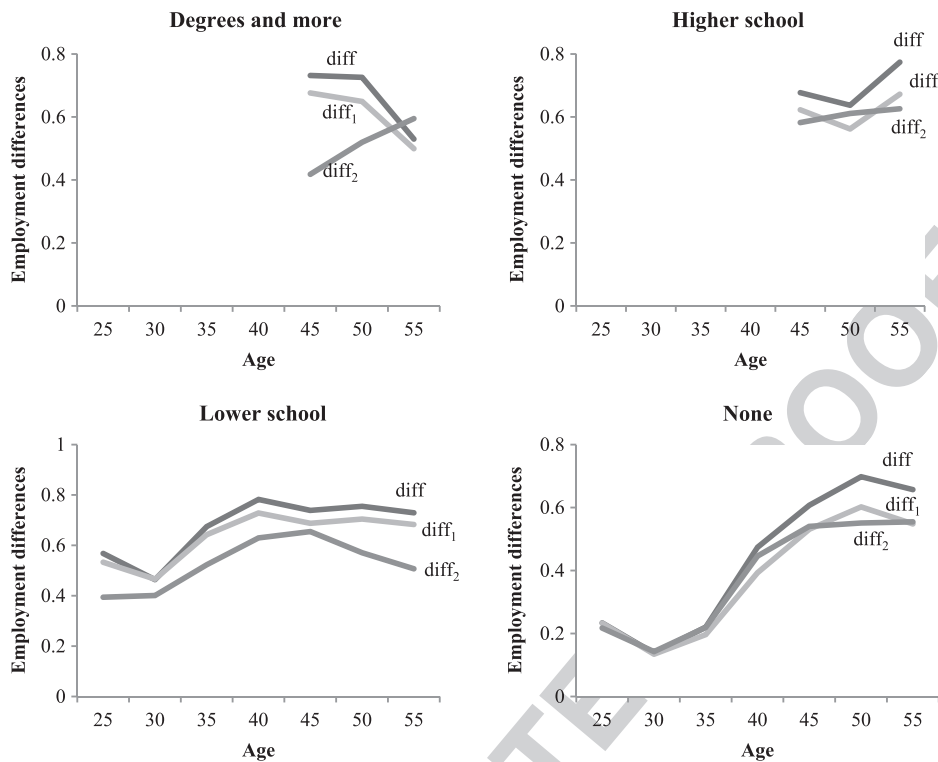
Note: Estimated difference in employment rates by disability status (diff), reported difference in employment rates by disability status for measure 1 (diff₁), reported difference in employment rates by disability status for measure 2 (diff₂)

Figure 4. Difference in employment rates by disability status based on reported and estimated disability—women

If the misclassification error was purely random, then it would generate an attenuation bias on the difference in employment by disability status. This is what we actually observe in our sample as in both Figures 3 and 4, *diff* lies always above $diff_1$ and $diff_2$. If false positives arose as justification for non-employment, then the bias could have been in the opposite direction. However, we have found that the proportion of false positives is very low. If false negatives are disproportionally not in employment, then they will contribute to the underestimated impact of disability on employment. The bias is greater using the second measure, $diff_2 < diff_1 < diff$. This arises despite the fact that the proportion of false negatives for measure 1 is larger. The reason is that, as we described earlier and show in the APPENDIX, the bias depends more on the false positives when d is less than 0.5.

The effect of misclassification on employment estimates does not vary significantly with gender. From the graphs, it seems that the estimated difference in employment by disability is more affected by the misclassification error for men than for women.

For men, the bias is quite small for those with higher school or no qualifications. On the other hand, it is quite large for those with lower school qualifications or degrees and above; in the former group, it should be noted that the biggest bias is observed for young men younger than 35 years old. The fact that the bias is so big for the two groups (lower school qualifications and degrees and above) can be understood by the very small false positive rates (qs) in these groups compared with the respective rates in the other qualification groups. It is worth mentioning that in the lower school qualifications group, the difference between the employment rates obtained from estimated and reported disability is increasing with age, which is consistent with the increasing misclassification error for this group.

The picture is much better for women independently of qualifications. The bias is tiny for women with lower school qualifications in the first measure across all age groups. The biggest impact of misclassification error on employment estimates is observed for women with degrees and above at the age of 45 years. We should highlight the fact that there seem to be no differences between the actual and reported employment rates for women with no qualifications younger the age of 40 years.

5. CONCLUSION

This paper looks at misclassification error in the reporting of disability status. There are three parameters of interest: (1) the misclassification probability and how it differs across sociodemographic groups; (2) the prevalence of disability; and (3) the difference in employment rates by disability status. The paper shows that all these parameters can be identified if

(a)   there are two 'noisy' indicators of the same latent disability construct that are observed for the same sample of individuals;
(b)   the researcher is prepared to assume that the misclassification probabilities are independent, for example, that over reporting on one measure does not increase the chance of over reporting on the other;
(c)   the expected difference in employment by disability status is not zero; and
(d)   the misreporting probability is independent of employment.

These assumptions are weaker than those invoked in some other studies that attempt to correct for misclassification errors in measured disability. For example, it is often assumed that the probability of over reporting is equal to the probability of under reporting and that these are both less than 50%. We are able to reject these restrictions. We find that the false positive rate is tiny compared with the probability that a given disabled person 'hides' his status. Of course, the plausibility and the validity of such assumptions must be specific to the actual datasets and the precise questions being asked, so we make no claim to be able to generalize our results to other studies of disability. Our method is ideally suited to invoking a slightly different set of assumptions.

We find that the over reporting rate is very small. Disability rates based on a direct question on whether you are disabled or not are quite close to the rate estimated from the data on the basis of the stated assumptions. The difference in employment by disability status is large and is slightly understated when derived directly from observed employment by reported disability.

Clearly some will disagree with the assumptions made in this paper, and thus, a possible future avenue for research will be to develop a methodology allowing this to be weaker. One possible technique would be to exploit restrictions on the unknown parameters across X, assuming, for example, that rates of false positives and negatives are the same for men and women. Another extension would be to think of each measure picking up different levels of severity of the disability, thus relaxing the assumption that the underlying *unobserved* variable is the same. In essence, this would be some kind of ordered logit or probit model where the thresholds are allowed to differ across individuals in some structured way.

APPENDIX: BIAS TO MEASURES OF EMPLOYMENT

We can write the difference in employment between those with $R_1 = 1$ and $R_1 = 0$ as

$$\Pr(E = 1 | R_1 = 1) - \Pr(E = 1 | R_1 = 0) = (e_D - e_{ND})\left(\frac{d(1 - p_1)}{d(1 - p_1) + (1 - d)q_1} - \frac{dp_1}{dp_1 + (1 - d)(1 - q_1)}\right).$$

We can then derive the alternative coefficient

$$1 - \frac{\Pr(E = 1|R_1 = 1) - \Pr(E = 1|R_1 = 0)}{(e_D - e_{ND})}$$

as

$$\frac{(1-d)q_1}{d(1-p_1) + (1-d)q_1} + \frac{dp_1}{dp_1 + (1-d)(1-q_1)}$$

Differentiating this w.r.t. p_1 gives   Q5

$$\frac{(1-d)q_1}{d(1-p_1) + (1-d)q_1}\frac{d}{d(1-p_1) + (1-d)q_1} + \frac{(1-d)}{dp_1 + (1-d)(1-q_1)}\frac{d(1-q_1)}{dp_1 + (1-d)(1-q_1)} \quad (A1)$$

and w.r.t. q_1 gives

$$\frac{(1-d)}{d(1-p_1) + (1-d)q_1}\frac{d(1-p_1)}{d(1-p_1) + (1-d)q_1} + \frac{dp_1}{dp_1 + (1-d)(1-q_1)}\frac{(1-d)}{dp_1 + (1-d)(1-q_1)} \quad (A2)$$

Simplifying the aforementioned equation and setting equation (A1) less than equation (A2), we can then   Q6
derive the conditions under which differences in q_1 have a bigger effect on the attenuation coefficient than
p_1. These simplify down to

$$(1 - q_1 - p_1) < \left[\frac{1 - \Pr(R_1 = 1)}{\Pr(R_1 = 1)}\right]^2 (1 - p_1 - q_1).$$

Thus, if $p_1 + q_1 < 1$, then q_1 will have a bigger effect on the attenuation coefficient (the bias in the estimated
employment differences by disability status) than p_1 when $\Pr(R_1 = 1)$ is less than 0.5. Likewise for measure 2.

ACKNOWLEDGEMENTS

REFERENCES

Aigner DJ. 1973. Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics* **2**:   Q7
49–59.
Baldwin M, Schumacher E. 2002. A note on job mobility among workers with disabilities. *Industrial Relations* **41**: 430–441.   Q8
Barron B. 1977. The effects of misclassification on the estimation of relative risk. *Journal of Biometrics* **33**: 414–418.   Q9
Berzofsky M, Biermer P, Kalsbeek W. 2008. A brief history of classification error models. Joint statistical meetings,   Q10
American Statistical Association.
Bollinger C. 1996. Bounding the effects of measurement error in regressions involving dichotomous variables. *Journal of
Econometrics* **73**: 387–399.
Bollinger C. 2003. Measurement error in human capital and the black-white wage gap. *The Review of Econometrics and*   Q11
Statistics **85**: 578–585.

Bound J. 1989. The health and earnings of rejected disability insurance applicants. *The American Economic Review* **79**: 482–503.

Bound J. 1991. Self-reported versus objective measures of health in retirement models. *Journal of Human Resources* **26**: 106–138.

Bound J, Burkhauser R. 1999. Economic analysis of transfer programs targeted on people with disabilities. *Handbook of Econometrics* **5**: 3417–3528.

Bound J, Brown C, Mathiowetz N. 2001. Chapter 59: measurement error in survey data. *Handbook of Labour Economics* Q12 **3C**: 3705–3843.

Bowe F. 1993. Statistics, politics, and employment of people with disabilities. *Journal of Disability Policy Studies* **4**: 84–91.

Bross I. 1954. Misclassification in 2x2 tables. *Biometrics* **10**: 478–486.

Diamond E, Lilienfield A. 1962. Misclassification errors in 2x2 tables with one margin fixed: some further comments. Q13 *American Journal of Public Health and the Nations Health* **52**: 2106–2110.

Goldberg J. 1975. The effects of misclassification on the bias in the difference between two proportions and the relative Q14 odds in the fourfold table. *Journal of the American Statistical Association* **70**: 561–567.

Hale T. 2001. The lack of a disability measure in today's current population survey. *Monthly Labour Review* **124**: 37–58.

Höfler M. 2005. The effect of misclassification on the estimation of association: a review. *International Journal of Methods* Q15 *in Psychiatric Research* **14**: 92–101.

Horowitz J, Manski CF. 1995. Identification and robustness with contaminated and corrupted data. *Econometrica* **63**: 281–302.

Hui SL, Walter SD. 1980. Estimating the error rates of diagnostic tests. *Biometrics* **36**: 167–171.

Kapteyn A, Smith JP, van Soest AHO. 2007. Vignettes and self-reports of work disability in the United States and the Q16 Netherlands. *American Economic Review* **97**: 461–473.

Kerkhofs M, Lindeboom M. 1995. Subjective health measures and state dependent reporting errors. *Health Economics* **4**: 221–235.

Keys A, Kihlberg J. 1963. Effect of misclassification on estimated relative prevalence of a characteristic: Part I. Two pop- Q17 ulations infallibly distinguished. Part II. Errors in two variables. *American Journal of Public Health and the Nations Health* **53**: 1656–1665.

Klepper S. 1984. Bounding mean regressions when a binary regressor is mismeasured. *Journal of Econometrics* **37**: 343–359. Q18

Kreider B. 1999. Latent work disability and reporting bias. *Journal of Human Resources* **34**: 734–769.

Kreider B. 2006. Partially identifying the prevalence of health insurance given contaminated sampling response error. Q19 *Working Paper* Department of Economics: Iowa State University.

Kreider B, Pepper J. 2007. Disability and employment: revaluating the evidence in light of reporting errors. *Journal of the American Statistical Association* **102**: 432–441.

Kreider B, Pepper J. 2008. Inferring disability status from corrupt data. *Journal of Applied Econometrics* **23**: 329–349.

Lessler J, Kalsbeek W. 1992. *Nonsampling Error in Surveys*. Wiley: New York. Q20

Lindeboom M, Kerkhofs M. 2004. Subjective health measures, reporting errors and endogeneity in the relationship between Q21 health and work. CEIS Research Paper 46.

Lindeboom M, Kerkhofs M. 2009. Subjective health measures, reporting errors and endogeneity in the relationship between Q22 health and work. *Journal of Applied Econometrics* **24**: 1024–1046.

Lohr SL. 1999. *Sampling: Design and Analysis*. Houghton Mifflin: Boston.

Molinary F. 2008. Partial identification of probability distributions with misclassified data. *Journal of Econometrics* **144**: Q23 81–117.

Morrissey MJ, Spiegelman D. 1999. Matrix methods for estimating odds ratios with misclassified exposure data: extensions Q24 and comparisons. *Biometrics* **55**: 338–344.

Myers R. 1982. Why do people retire from work early? *Social Security Bulletin* **45**: 1925–1928.

Newell D. 1962. Errors in the interpretation of errors in epidemiology. *American Journal of Public Health and Nations* Q25 *Health* **52**: 1925–1928.

O'Donnell O 1998. The effect of disability on employment allowing for work incapacity. *Studies in Economics 9813* Q26 Department of Economics: University of Kent.

Phelps ES. 1972. The statistical theory of racism and sexism. *American Economic Review* **62**: 659–661. Q27

Savoca E. 2000. Measurement errors in binary regressors: an application to measuring the effects of specific psychiatric Q28 diseases on earnings. *Health Services and Outcomes Research Methodology* **1**: 149–164.

Tenenbein A. 1972. A double sampling scheme for estimating from misclassified multinomial data with application to sampling inspection. *Technometrics* **14**: 187–202.

The Equality and Human Rights Comission (EHRC). 2010. Equality Act 2010 **1**: Chapter 15. Q29

Wallace R, Herzog A. 1995. Overview of the health measures in the health and retirement study. *The Journal of Human* Q30 *Resources* **30**: S84–S107.

Author Query Form

Dear Author,

During the copyediting of your paper, the following queries arose. Please respond to these by annotating your proofs with the necessary changes/additions.
- If you intend to annotate your proof electronically, please refer to the E-annotation guidelines.
- If you intend to annotate your proof by means of hard-copy mark-up, please refer to the proof mark-up symbols guidelines. If manually writing corrections on your proof and returning it by fax, do not write too close to the edge of the paper. Please remember that illegible mark-ups may delay publication.

Whether you opt for hard-copy or electronic annotation of your proofs, we recommend that you provide additional clarification of answers to queries by entering your answers on the query sheet, in addition to the text mark-up.

Query No.	Query	Remark
Q1	AUTHOR: "Kerkhofs and Lindeboom, 2009" is cited in text but not given in the reference list. Please provide details in the list or delete the citation from the text.	
Q2	AUTHOR: Please check the numbering of the ff. equations: A.1, A.2.....etc.	
Q3	AUTHOR: "…variables using IV." The meaning of this is not clear. Please reword or confirm that this is correct.	
Q4	AUTHOR: The meaning of "above" is not clear. Please consider changing it to "aforementioned…" to avoid confusion.	
Q5	AUTHOR: Please define w.r.t.	
Q6	AUTHOR: Numbering for equations in the Appendix were changed. Please check.	
Q7	AUTHOR: Reference "Aigner (1973)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q8	AUTHOR: Reference "Baldwin & Schumacher (2002)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q9	AUTHOR: Reference "Barron (1977)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	

| Query No.	Query	Remark
Q10	AUTHOR: Reference "Berzofsky et al. (2008)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q11	AUTHOR: Reference "Bollinger (2003)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q12	AUTHOR: Reference "Bound et al. (2001)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q13	AUTHOR: Reference "Diamond & Lilienfield (1962)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q14	AUTHOR: Reference "Goldberg (1975)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q15	AUTHOR: Reference "Höfler (2005)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q16	AUTHOR: Reference "Kapteyn et al. (2007)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q17	AUTHOR: Reference "Keys & Kihlberg (1963)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q18	AUTHOR: Reference "Klepper (1984)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q19	AUTHOR: Reference "Kreider (2006)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q20	AUTHOR: Reference "Lessler & Kalsbeek (1992)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q21	AUTHOR: Reference "Lindeboom & Kerkhofs (2004)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q22	AUTHOR: Reference "Lindeboom & Kerkhofs (2009)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q23	AUTHOR: Reference "Molinary (2008)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q24	AUTHOR: Reference "Morrissey & Spiegelman (1999)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	

Query No.	Query	Remark
Q25	AUTHOR: Reference "Newell (1962)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q26	AUTHOR: Reference "O'Donnell (1998)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q27	AUTHOR: Reference "Phelps (1972)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q28	AUTHOR: Reference "Savoca (2000)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q29	AUTHOR: Reference "The Equality and Human Rights Comission (EHRC) (2010)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q30	AUTHOR: Reference "Wallace & Herzog (1995)" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	

# USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-Annotate PDFs: <u>Adobe Acrobat Professional</u> or <u>Adobe Reader</u> (version 7.0 or above). (Note that this document uses screenshots from <u>Adobe Reader X</u>)
The latest version of Acrobat Reader can be downloaded for free at: <u>http://get.adobe.com/uk/reader/</u>

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:

| | | | | 1 / 27 | | | 70.4% | | | | | Tools | Comment | Share |
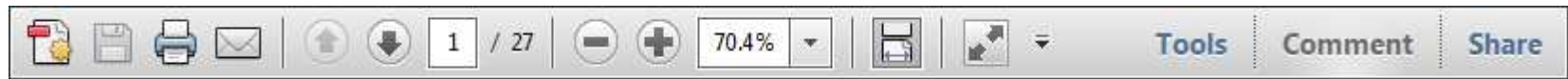
This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We've picked out some of these tools below:
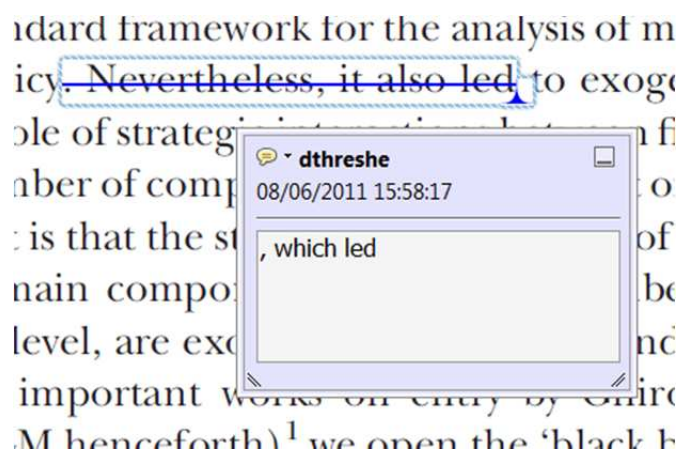
▼ Annotations

---
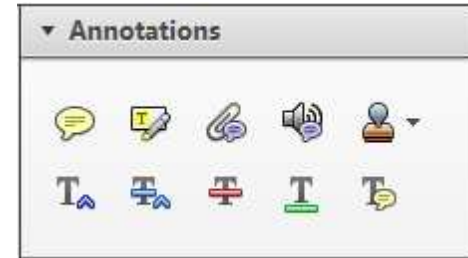
1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it

- Highlight a word or sentence.
- Click on the Replace (Ins) icon in the Annotations section.
- Type the replacement text into the blue box that appears.

ndard framework for the analysis of m
icy. Nevertheless, it also led to exoge
ole of strategi...
nber of comp...
: is that the st...
nain compo...
level, are exc...
important wo...
M henceforth)[1] we open the 'black
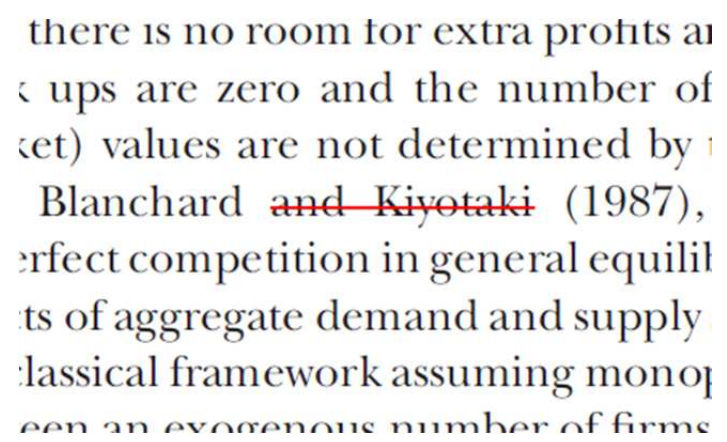
dthreshe
08/06/2011 15:58:17
, which led

---

2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it

- Highlight a word or sentence.
- Click on the Strikethrough (Del) icon in the Annotations section.

there is no room for extra profits al
ups are zero and the number of
et) values are not determined by
Blanchard and Kiyotaki (1987),
rfect competition in general equili
ts of aggregate demand and supply
lassical framework assuming mono
een an exogenous number of firms

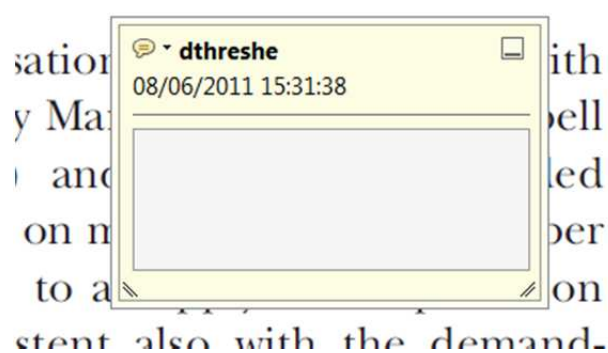### 3. Add note to text Tool – for highlighting a section to be changed to bold or italic.

Highlights text in yellow and opens up a text box where comments can be entered.

How to use it

- Highlight the relevant section of text.
- Click on the Add note to text icon in the Annotations section.
- Type instruction on what should be changed regarding the text into the yellow box that appears.

namic responses of mark ups
ent with the VAR evidence

sation...
y Ma...
and...
on n...
to a...
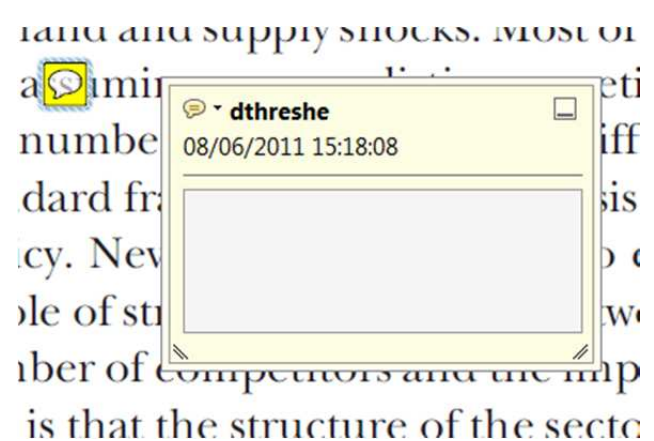stent also with the demand-

dthreshe
08/06/2011 15:31:38

---

4. Add sticky note Tool – for making notes at specific points in the text.

Marks a point in the proof where a comment needs to be highlighted.

How to use it

- Click on the Add sticky note icon in the Annotations section.
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the yellow box that appears.

land and supply shocks. Most of
a...mi...
numbe...
dard fr...
cy. New...
ole of st...
ber of competitors and the imp
is that the structure of the secto

dthreshe
08/06/2011 15:18:08

# USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate pace in the text.

How to use it

- Click on the Attach File icon in the Annotations section.
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.

END

0.20
0.15
0.10
0.05
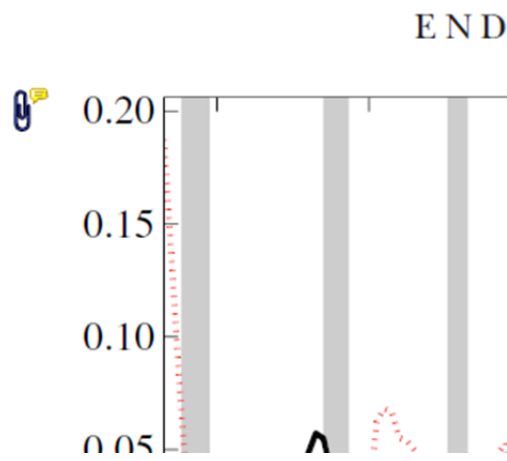
6. Add stamp Tool – for approving a proof if no corrections are required.

Inserts a selected stamp onto an appropriate place in the proof.

How to use it

- Click on the Add stamp icon in the Annotations section.
- Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears).
- Click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).

or the business cycle, starting with the
on perfect competition, constant ret
production. In this environment goods
extra profits and the structure of market
he APPROVED p
etermined by the model. The New-Key
otaki (1987), has introduced produc
general equilibrium models with nomina
d and supply shocks. Most of this literat


Drawing Markups

7. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.

Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks..

How to use it

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
- Double click on the shape and type any text in the red box that appears.

dthreshe
08/06/2011 16:23:54

Move to the bottom of the left column of panels.
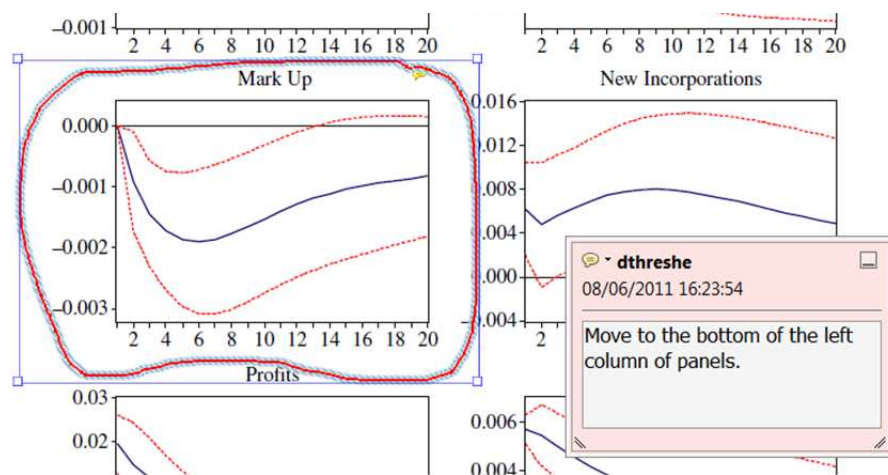
For further information on how to annotate proofs, click on the Help menu to reveal a list of further options:

ecoj_2384_CrxRev2_EV_19-Jul-10.pdf - Adobe Reader

File   Edit   View   Window   Help

? Adobe Reader X Help...   F1

About Adobe Reader X...
About Adobe Plug-Ins...

Improvement Program Options...

Digital Editions

Online Support...
Repair Adobe Reader Installation
Check for Updates...

Purchase Adobe Acrobat

Tools   Comment   Share

Annotations

Drawing Markups

Comments List (14)

Find

dthreshe