



Scand J Work Environ Health 2014;40(5):493-501

<https://doi.org/10.5271/sjweh.3440>

Published online: 11 Jun 2014, Issue date: 01 Sep 2014

Assessing work ability - a cross-sectional study of interrater agreement between disability claimants, treating physicians, and medical experts

by [Dell-Kuster S](#), [Lauper S](#), [Koehler J](#), [Zwimpfer J](#), [Altermatt B](#), [Zwimpfer T](#), [Zwimpfer L](#), [Young J](#), [Bucher HC](#), [Nordmann AJ](#)

This cross-sectional study quantifies disagreement in assessment of work ability of disability claimants referred to a multidisciplinary assessment center. The high level of disagreement calls for a careful evaluation of the disability assessment process in an effort to reduce the disagreement between expert teams, treating physicians, and claimants.

Affiliation: Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Hebelstrasse 10, CH - 4051 Basel, Switzerland. Alain.Nordmann@usb.ch

The following article refers to this text: [2015;41\(1\):102-104](#)

Key terms: [claimant](#); [cross-sectional study](#); [disability](#); [disability claimant](#); [disability insurance claim](#); [disability pension](#); [disagreement](#); [interrater agreement](#); [medical expert](#); [percentage work ability](#); [physician](#); [work ability](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/24919161



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

Assessing work ability – a cross-sectional study of interrater agreement between disability claimants, treating physicians, and medical experts

by Salome Dell-Kuster, MD, MSc,¹ Simon Lauper, MD,² Joerg Koehler, MSc,² Johanna Zwimpfer, MD,² Benedikt Altermatt, BSc,² Tibor Zwimpfer, BSc,² Leon Zwimpfer, BSc,² Jim Young, PhD,¹ Heiner C Bucher, MD, MPH,¹ Alain J Nordmann, MD, MSc¹

Dell-Kuster S, Lauper S, Koehler J, Zwimpfer J, Altermatt B, Zwimpfer T, Zwimpfer L, Young J, Bucher HC, Nordmann AJ. Assessing work ability – a cross-sectional study of interrater agreement between disability claimants, treating physicians and medical experts. *Scand J Work Environ Health*. 2014;40(5):493–501. doi:10.5271/sjweh.3440

Objectives It is unclear to what extent assessments of work ability differ between disability claimants, their treating physicians, and multidisciplinary medical expert teams.

Methods We compared assessments of work ability for consecutive disability claimants referred to a multidisciplinary assessment center in Switzerland over a 4-year period. Assessments were made for the last job (LJ) prior to claiming a disability benefit and an alternative job (AJ) thought to suit the claimant's physical and mental abilities. Mean differences (MD) in percentage work ability between assessments from claimants, physicians, and experts were then estimated in a linear regression model.

Results The 3562 claims made during the study period were mostly due to musculoskeletal and depressive disorders. Assessments differed little between claimants and physicians [LJ MD 1.3% (95% confidence interval [95% CI] 0.5–2.2%); AJ MD 11% (95% CI 10–12%)]. Experts on average assessed a claimant's work ability higher than either the claimant or physician, particularly in the AJ [MD between expert and claimant 57% (95% CI 56–58%) and between expert and physician 46% (95% CI 45–48%)].

Conclusions Assessments of work ability differed substantially between experts in multidisciplinary medical teams and both claimants and their treating physicians. A careful evaluation of the disability assessment process is needed in an effort to reduce disagreement between expert teams and treating physicians and so improve acceptance of the process.

Key terms disability insurance claim; disability pension; disagreement; percentage work ability.

Public spending on disability benefits represents, on average, 2% of gross domestic product across member countries of the Organisation for Economic Cooperation and Development (OECD) (1). In OECD countries, around 6% of the working-age population rely on disability benefits, with up to 10–12% on such benefits in some northern European countries. In Switzerland, the number of disability benefits recipients increased by 27% from 2000 to 2005 (2). The deficit in Swiss disability insurance increased accordingly and led to a 5th revision of the Swiss disability insurance in 2008. This revision had reintegration of disability claimants as its primary goal, and disability benefits were restricted. The number of disability benefit recipients then decreased

by 7% between 2005 and 2012 (2). In 2012, 5% of the insured workforce in Switzerland were receiving a disability benefit (2).

Criteria and procedures for assessing work ability and eligibility for disability benefits differ substantially between countries. In many countries, treating physicians play a pivotal role in this process. In Switzerland, claimants for a disability benefit apply to cantonal offices of the Swiss disability insurance. Disability insurance officers then request medical reports from the treating physicians, together with an assessment about the claimant's work ability in their last job (LJ) and an alternative job (AJ) thought to be suitable considering the claimant's physical and mental abilities. The treat-

¹ Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland.

² Aerztliches Begutachtungsinstitut ABI GmbH, Basel, Switzerland.

ing physician is not asked to name a specific AJ but to describe the characteristics of such a job (eg, no lifting of objects weighing >10 kilos, able to change from standing to sitting, able to have a short break after any full hour of work, etc.). If the underlying diagnoses seem inconsistent with these assessments, disability insurance officers can refer a claimant to medical experts at a multidisciplinary assessment center (eg, if a primary care physician reports a diagnosis of osteoarthritis of the knee and certifies full disability even for a sitting job). The number of disability claimants referred to assessment centers is about 10% of all claims (3).

Assessing work ability is a complex process with many sources of variation. Studies have focused on variation between physicians (4, 5) or between physicians and claimants (6, 7). Different means of workplace assessment or different levels of information about the medical condition and psychosocial factors seem to account for most of the variation between physicians; individual characteristics and opinions seemed to have little influence (4). Variation arising from the collection of information, interpretation, and documentation of the results could be reduced by adopting transparent and standardized processes (5). Patients usually assess their own work ability lower than their physicians do (7, 8). Patients with somatoform (7) or depressive disorder (6, 9) seem to give more variable assessments than other patients. The reason for the variation in work ability assessment among patients with these diagnoses could be that such diagnoses are not easily ascertainable. However, the financial impact is large since these diagnoses are very common.

The reimbursement of experts by the Swiss disability insurance creates a conflict of interest for experts. Recently, lawyers of disability claimants and the media have criticized the quality of experts' reports (10). The level of disagreement between the assessments of treating physicians and multidisciplinary medical expert teams could be considerable but has never been quantified. In this study, we estimated differences between assessments of work ability for claimants referred to a multidisciplinary assessment center in the northwest of Switzerland.

Methods

The study protocol for this retrospective analysis of routinely collected data was submitted to the Federal Commission of Experts for Professional Secrecy in Medical Research, which exempted the study from formal ethical committee approval. The study follows guidelines for reporting reliability and agreement studies (GRRAS) (11).

Data collection

We analyzed data for all consecutive disability insurance claimants referred to a single multidisciplinary assessment center (Aerztliches Begutachtungsinstitut) in Basel, Switzerland from January 2005 to December 2008. This center provides medical expertise for a variety of insurance organizations, primarily the Swiss disability insurance. We estimate that about 12% of all disability claims in Switzerland were handled by this assessment center during these four years.

Claimants' characteristics such as age, gender, nationality, and diagnoses with and without an influence on the working ability were recorded by a member of the multidisciplinary expert team, usually a specialist in general internal medicine. Diagnoses were routinely made based on history taking, clinical examinations, and laboratory analyses, and – where deemed necessary by the evaluating specialists – by additional examinations. Each specialist of the expert team coded diagnoses according to the International Classification of Diseases, version 10 (ICD-10). Somatoform disorders were only rated as having an influence on work ability when they were accompanied by either severe somatic or severe psychiatric comorbidity.

Work ability was assessed as a percentage, from 0% (a claimant unable to work at all) to 100% (a claimant fully able to work), and was then assigned to one of the disability categories used by the Swiss Accident Insurance Company: $\leq 30\%$ =severe restriction, 31–50%=moderate to severe restriction, 51–79%=mild to moderate restriction, 80–99%=mild restriction, and 100%=no restriction. A disability benefit will be granted only after a minimum of one year's sick-leave. The granted benefit is permanent if no changes in health status can be expected. Otherwise the recipient's health status is reevaluated every two to three years.

Physicians assessed a claimant's work ability using a standard form provided by the Swiss disability insurance approximately 6–12 months before the claimant was referred to the multidisciplinary assessment center. At the assessment center, after giving their medical history, claimants were asked to assess their own percentage work ability for their LJ and AJ suggested by their physician. Experts at the center were not blinded to these two earlier assessments. The expert team consisted of at least three medical specialists always including a specialist in general internal medicine and a psychiatrist. The experts' assessment of the claimants' work ability was reached by consensus after each had evaluated the claimant.

Statistical analysis

All data were stored anonymously. The anonymized data were handed over to independent statisticians for analy-

sis. All analyses were conducted using Intercooled Stata Version 11.2 for Macintosh (StataCorp, College Station, TX, USA). We report 95% confidence intervals (95% CI), rather than P-values, to emphasize clinical relevance over statistical significance because in this large data set, irrelevant differences are also statistically significant.

Categorical agreement

Raw categorical agreement (interrater agreement) was calculated as the number of exact categorical matches between two assessments divided by the total number of claimants (12). Observed and predicted probabilities of categorical work ability were found for both jobs and all three assessments, with predicted probabilities estimated in a proportional odds model (with interactions between assessments and jobs) and 95% CI based on a robust variance with a cluster for each claimant (13).

Analysis of the percentage work ability

Percentage work ability is shown in scatterplots for each pair of assessments. We estimated the mean difference (MD) between assessments with a linear regression model fit using generalized estimating equations and a cluster for each claimant (14). We also estimated the MD using a fractional logit model suitable for modeling responses bounded between zero and one which are often highly skewed (15). However, due to the large data set, estimates from both models were almost identical, and thus in the results section we only report MD from the linear model.

In two pre-specified subgroup analyses, we estimated the MD in percentage work ability assessments for claimants diagnosed with or without a severe depression (F32 and F33) thought to influence their work ability, or with or without any form of somatoform disorder (F45 and F68). These estimates were made using the full three-way interaction model with interactions between assessments, jobs, and subgroups. We expected that differences between expert and claimant and between expert and physician would be less among claimants diagnosed with severe depression (relative to those without severe depression), and greater among claimants with somatoform disorders (relative to those without such disorders).

Results

Over the study period, 3462 insurance claimants were referred to the assessment center: the median age of claimants was 47 years [interquartile range (IQR) 40–53] and 50% were female (table 1). One-third of the claimants were Swiss, one-third from countries in

the Balkans, and the last third from other countries. Musculoskeletal (group M) and psychiatric disorders (group F) were the most common diagnoses thought to influence work ability (table 2).

Categorical agreement

In general, treating physicians and claimants gave similar assessments with an overall agreement of 84% in the LJ and 68% in the AJ. However, typically experts assessed the claimant's work ability higher than the assessment of either the claimant or the physician,

Table 1. Claimant characteristics (N=3562). [IQR=interquartile range]

Characteristics	Median	IQR	N	%
Age (years)	47	40–53		
Gender				
Male			1797	50
Female			1765	50
Nationality				
Switzerland			1291	36
Countries in the Balkans ^a			1254	35
Southern Europe ^b			345	10
Turkey			368	10
Other Europe ^c			153	4
Other countries			151	4

^a Albania, Bosnia, Croatia, Kosovo, Macedonia, Montenegro, Serbia & Slovenia.

^b Spain, Portugal, Italy & Greece.

^c Austria, Belgium, Bulgaria, Czech Republic, Denmark, France, Great Britain, Germany, Hungary, Liechtenstein, The Netherlands, Poland & Romania.

Table 2. Claimant comorbidities. A maximum of five diagnoses per claimant was recorded (multiple diagnoses possible).

	N	%
Diagnoses with influence on work ability		
Dorsalgia (M54)	1698	48
Other dorsopathies, not elsewhere classified (M53)	1010	28
Depressive episode (F32)	816	23
Somatoform disorders (F45)	633	18
Recurrent depressive disorder (F33)	456	13
Shoulder lesions (M75)	295	8
Gonarthrosis (M17)	248	7
Other soft tissue disorders, not elsewhere classified (M79)	245	7
Other post-surgical states (Z98)	210	6
Other intervertebral disc disorders (M51)	195	6
Diagnoses without influence on work ability		
Obesity (E66)	1044	29
Mental and behavioral disorders due to use of tobacco (F17)	786	22
Psychological and behavioural factors associated with disorders or diseases classified elsewhere (F54)	673	19
Essential (primary) hypertension (I10)	669	19
Somatoform disorders (F45)	500	14
Disorders of lipoprotein metabolism and other lipidaemias (E78)	433	12
Non-insulin-dependent diabetes mellitus (E11)	352	10
Pain, not elsewhere classified (R52)	287	8
Dorsalgia (M54)	274	8
Other dorsopathies, not elsewhere classified (M53)	253	7

leading to a lower level of agreement (51% for both comparisons in the LJ). This effect was even more pronounced in the AJ, with a raw agreement of only 15% between experts and claimants and 20% between experts and physicians. The disagreement between the expert and the other assessors, particularly in the AJ, is illustrated by the different frequencies with which work ability categories were assigned (figure 1). There was little difference between the claimant-assigned LJ and AJ assessment frequencies; physicians assigned slightly higher levels of work ability in an AJ compared to a LJ while experts assigned markedly higher levels of work ability in an AJ.

The observed probabilities (from raw data) of being assessed in a certain work ability category closely matched the predicted probability of being in that cat-

egory under a proportional odds model (table 3). Odds ratios (OR) for this model imply higher probabilities of an expert assessing the claimant's work ability in a higher category compared to claimants or physicians (table 4). The difference between experts and claimants or physicians was even more pronounced in the AJ (with OR of 20 and 10 respectively) than in the LJ (with OR of 7 and 6 respectively).

Percentage work ability

Assessments of the percentage work ability in LJ and AJ are shown in scatterplots for each pair of assessments (overlying data jittered so each pair can be seen) with coincidental assessments on the diagonal line (figure 2). MD in percentage work ability were greater in the

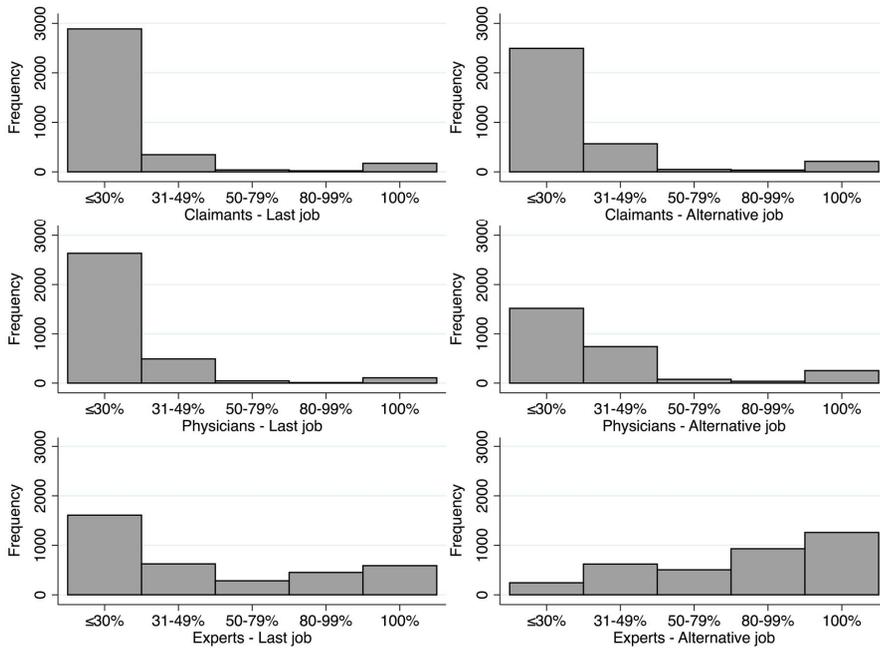


Figure 1. Frequency distribution of categorical assessment of ability to work. Histogram illustrating the frequencies of categories for work ability as assessed by the claimants, physicians, and experts. Left column assessments in the last job, the right column assessments in an alternative job.

Table 3. Percentage of observed (obs; from raw data) versus predicted (pred; from a proportional odds model) probabilities of categorical work ability as assessed by claimants, physicians and experts. [Obs=Observed probability; Pred=Predicted probability].

	N	N (missing)	Work ability categories									
			≤30%		31-50%		51-79%		80-99%		100%	
			Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
Last job												
Claimant	3462	100	83	83	10	11	1	2	1	2	5	2
Physician	3286	276	80	81	15	12	1	2	0.3	2	3	2
Expert	3556	6	45	40	18	28	8	8	13	10	17	14
Alternative job												
Claimant	3354	208	74	74	17	16	1	3	1	3	6	3
Physician	2622	940	58	59	28	23	3	5	1	6	10	7
Expert	3558	4	7	12	17	19	14	10	26	17	35	42

AJ than the LJ for every pair of assessments (table 4). The difference between the assessments of experts and claimants was highest, with a MD of 29% in the LJ (95% CI 28–31%) and 57% in the AJ (95% CI 56–58%). Assessments differed little between physicians and claimants, with a MD of 1.3% in the LJ (95% CI 0.5–2.2%) and 11% in the AJ (95% CI 10–12%). In our data, for LJ and AJ, only 5% and 6% of claimants, respectively, had a higher assessment from their physician than from the expert.

Sub-group analyses

Of 3305 claimants without missing records and with a diagnosis thought to influence their work ability, 370 (11%) suffered from a severe depressive disorder. In the LJ, the MD between experts and claimants in percentage work ability was similar among claimants with and without severe depression (21% among those with versus 26% among those without severe depression; table 5). However in the AJ, this MD between experts and claimants was reduced among those claimants with a severe depression [43% (95% CI 40–46%)] relative to those without such a depression [58% (95% CI 56–59%)].

Of 3561 claimants without missing records, 1123 (32%) had a somatoform disorder. In both jobs, the dif-

ference between experts and claimants in percentage work ability was greater among claimants with relative to those without somatoform disorder (in the LJ, 37% among those with versus 26% among those without somatoform disorder; in the AJ, 65% among those with versus 53% among those without somatoform disorder; table 5).

Missing assessments

Physicians did not assess work ability for 276 (8%) claimants in their LJ and for 940 (26%) claimants in the AJ. In the LJ, 100 (3%) claimants did not assess their own work ability; in the AJ, 208 (6%) claimants did not assess their own work ability. However, the distribution of expert assessments for these missing assessments was similar to the distribution of expert assessments where assessments of physicians and claimants were available (data not shown).

Discussion

In this study, we quantified disagreement in assessments of work ability for those disability claimants referred to a multidisciplinary assessment center. Assessments differed

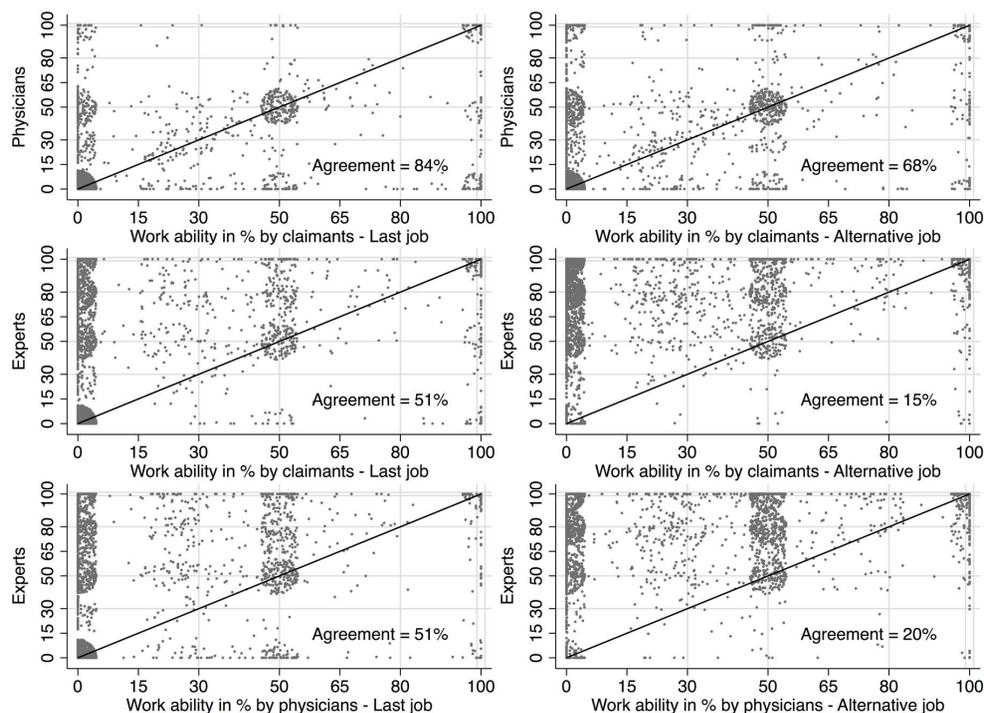


Figure 2. Scatterplot of the percentage work ability by each pair of assessments (overlying data jittered so each pair can be seen) with coincidental assessments on the diagonal line. Plots in the left column depict the assessments in the last job, and those in the right column an alternative job. Observations above the line indicate the assessments on the vertical axis is higher than the assessment on the horizontal axis; observations below the line indicate the reverse. Fine grey lines denote the disability categories used by the Swiss Accident Insurance Company: $\leq 30\%$ =severe restriction, 31–50%=moderate-to-severe restriction, 51–79%=mild to moderate restriction, 80–99%=mild restriction, 100%=no restriction.

little between claimants and their treating physicians, both in their LJ and in an AJ. In contrast, multidisciplinary expert teams typically assessed a claimant as having a higher work ability, particularly in an AJ. This disagreement between experts, physicians, and claimants was smaller among claimants diagnosed with severe depression but greater among those with somatoform disorders.

Strengths

The strength of our study is the consecutive inclusion of a large number of claimants referred to a single center over a 4-year period. Claimants referred to this center appear similar to Swiss disability claimants in general, in that one third of those granted a disability benefit are Swiss and two thirds are foreign nationalities and have a similar age and gender distribution to our claimants (16).

Table 4. Comparison of categorical and percentage work ability. Odds ratios (OR) under a proportional odds model with an interaction terms between all assessors and both jobs with robust variance estimator allowing for clustering within claimants (first two columns); mean differences (MD) in percentage work ability from a linear regression model using generalized estimating equations with an unstructured correlation matrix (last two columns) (N=3562 claims). [95% CI=95% confidence interval]

Comparison	OR ^a	95% CI	MD %	95% CI
Last job				
Physician versus claimant	1.19	1.09–1.30	1	0.5–2.2
Expert versus claimant	7.38	6.68–8.15	29	28–31
Expert versus physician	6.20	5.65–6.80	28	26–29
Alternative job				
Physician versus claimant	2.01	1.84–2.19	11	10–12
Expert versus claimant	20.4	18.5–22.4	57	56–58
Expert versus physician	10.1	9.29–11.1	46	45–48
Alternative versus last job (patients)	1.70	1.57–1.84	6	5.1–6.9
Alternative versus last job (physicians)	2.87	2.64–3.11	16	14–17
Alternative versus last job (experts)	4.70	4.37–5.05	34	33–35

^a OR >1 implies that the first rater is more likely than the second rater to provide a higher assessment of a claimant's ability to work.

There were few missing assessments of work ability in the LJ. There were considerably more missing physician assessments of work ability in an AJ, but the distribution of expert assessments for these claimants was similar to the distribution where assessments were available from the physician. Thus the validity of our study does not seem to be threatened by missing assessments.

Limitations

Our study has several limitations. The experts were not blinded to the assessments of the physician and claimant. It is unclear, however, whether this introduced a bias and if so, whether expert assessment was higher or lower than it would be otherwise. All three assessments were not made at the same point in time. The claimant's physician made the first assessment, approximately 6–12 months before the claimant and experts made their assessments. The relatively high agreement between the assessments of claimant and their physician suggests that the delay between assessments did not materially affect our results.

Explanations for disagreement in the work ability assessments

A number of factors may contribute to the difference between experts', physicians', and claimants' assessment of the disability claimants' work ability. First, the claimants included in our study represent a selected sample of all disability claimants. These were selected because an insurance officer thought the physicians' assessment of work ability was unreliable. This selection was reflected in a high number of diagnoses of depressive and somatoform disorders. It is not clear whether the results of our single center study can be generalized to other assessment centers in Switzerland or other countries with different disability assessment procedures. In a recent Dutch study, there was a high agreement between disability claimants' expectations of a disability benefit

Table 5. Subgroup analyses: disagreement in work ability (between experts, physicians, and claimants) among patients with severe depression or somatoform disorders as compared to those without such diagnoses [95% CI=95% confidence interval; MD=mean difference].

	Without severe depression		With severe depression ^a		Difference between subgroups		Without somatoform disorder		With somatoform disorder ^b		Difference between subgroups	
	MD %	95% CI	MD %	95% CI	MD %	95% CI	MD %	95% CI	MD %	95% CI	MD %	95% CI
Last job												
Expert versus claimant	26	25–28	21	18–24	-5	-8.7–-1.8	26	24–27	37	34–39	11	8.4–14
Expert versus physician	25	23–26	21	17–24	-4	-7.6–-0.3	25	23–26	34	32–37	9	6.3–12
Alternative job												
Expert versus claimant	58	56–59	43	40–46	-14	-17–-11	53	52–55	65	63–67	12	9.3–14
Expert versus physician	46	45–48	35	32–39	-11	-15–-6.6	43	41–45	53	50–55	10	6.7–12

^a Patients with severe depression: 370 (11% of 3305 without missing records - only diagnosis with influence on work ability considered).

^b Patients with somatoform disorder: 1123 (32% of 3561 without missing records - diagnoses with and without influence on work ability considered).

and the subsequent receipt of such a benefit (17), which seems at odds with our results. However, the Dutch study was of unselected disability claimants whereas the claimants in our study were selected because their assessments were considered unreliable.

Our data suggests that the difference between assessments varies with the diagnosis, with greater agreement when claimants have severe depression and less agreement when claimants have somatoform disorders. This was expected – our view is that a diagnosis of depression is an implicit recognition that a patient will find it difficult to work in any capacity, whereas it is not clear what a diagnosis of a somatoform disorder should imply about a patient's work ability. It seems likely that agreement between physicians and experts would be substantially higher when assessing claimants with diagnoses based on objective findings such as disabling stroke, metastatic cancer or chronic obstructive lung disease. However, we were unable to show this with these data because such claimants are seldom referred to a multidisciplinary assessment center.

Second, when assessing a patient's work ability in either job, physicians may be reluctant to disagree with their patient because of the (sub) conscious fear of losing the patient to another physician. In a recent survey, only 7% of general practitioners (GP) in Sweden and 18% in Norway said they were worried about losing patients if they did not provide a sick-leave certificate (18). However, patients seem to have a strong influence on such assessments: in one Swedish medical audit, a sick-leave certificate was issued in 87% of cases even when the GP did not think it was warranted (19).

Third, there is the potential for conflicting interests to contribute to the disagreement between experts and physicians. Swiss Federal law requires that an independent interdisciplinary expert team assess disability claims to mitigate bias but multidisciplinary assessment centers are paid by disability insurers and not by a neutral third organization. We cannot rule out unconscious bias by experts.

Fourth, expert assessment does not necessarily guarantee high quality and reproducible assessment of a claimant's work ability. In a pilot study of expert medical assessments submitted to Swiss accident, disability, and liability insurance agencies in 2008, 20% of a random sample of 97 assessments were found to be of insufficient quality (20). In a US study, there was substantial variability in disability assessments when 48 identical clinical vignettes were presented to 36 GP experienced in disability assessments (21).

Lastly, claimants and physicians may place more weight on environmental and personal factors than experts. This could partly explain the similarity of assessments by physicians between LJ and AJ, as these factors will influence work ability in both jobs. In inter-

views with a random sample of 60 insurance physicians from the Dutch National Institute for Employee Benefit Schemes, when determining work ability, experts tended to consider different domains (like "functions and structures" when claimants had musculoskeletal disorders and "participation" when claimants had psychiatric and other disorders) as main drivers for disability judgments for determining work ability (22). "Environmental factors" (eg, workplace factors, conflicts with employer or family) and "personal factors" (eg, coping, compliance to therapy, illness behavior, motivation, age) were seldom considered when determining work ability. Whether environmental and personal factors should be considered in determining work ability – and how much influence they should have – is open to debate. In 2004, the Swiss Federal Court of Justice ruled that, while a diagnosis of somatoform pain disorder does not of itself qualify a claimant for a disability benefit, there might be exceptions to this rule where patients cannot be expected to wilfully overcome their pain due to the presence of specific coercive comorbidities or circumstances (23). In accordance with this ruling, psychosocial factors per se are generally not considered by medical expert teams when assessing a claimant's work ability.

Implications and further research

Sickness certification and assessment of work ability are difficult tasks. In a survey of all physicians in Sweden, 60% found it difficult to assess a patient's capacity to work and give a prognosis for the duration of a patient's incapacity (19). The same is likely to be true for disability assessments: most physicians do not receive any training in determining work ability (24) and it seems sensible to give medical students some training in sick-leave certification and disability assessment. But even experienced disability assessors will vary in their assessments, although only small to moderate variation would be expected (4).

Unfortunately there is a lack of well-validated tools or procedures to help physicians assess disability (25). In the last decade, the International Classification of Functioning, Disability and Health (ICF) has been used to improve the assessment of disability, but even the use of this classification does not necessarily result in a reliable standardized evaluation of disability (26). Thus, there is an urgent need for validated instruments with which to assess work ability. It has been proposed that standardizing information collection – using clear guidelines and a reliable and validated instrument to document the assessment – and using multiple and specially trained assessors may reduce variation in work disability assessment (5). While these suggestions clearly make sense, studies to prove the effectiveness of their application are still lacking.

The high level of disagreement between physicians and experts when assessing work ability has important consequences. Controversial disability requests will generally only be granted if substantiated by experts, so physicians potentially harm both the disability claimant and society by certifying partial or full disability in cases where experts subsequently fail to confirm it. A disability claimant, backed by a physician's partial or full disability certificate, may be less willing to reintegrate into the workforce after a negative disability pension decision. These disability claimants will frequently have lost their job during the assessment process and it may not be easy to find another. Further studies should investigate the reasons for the high level of disagreement between physicians and multidisciplinary expert teams when assessing a disability claimant's work ability. Once reasons have been identified, they may be tackled by specific interventions.

Concluding remarks

The process of assessing a disability claimant's work ability should be improved by measures that narrow the wide gap between assessments. To achieve this goal, primary-care physicians need training about criteria and standards used in multidisciplinary assessment centers and the law under which experts operate. For medical expert teams, there is a need to develop standardized (and if possible validated) criteria with which to reproducibly assess work ability among disability claimants. The development, implementation, and careful evaluation of these criteria should reduce disagreement when assessing a disability claimants' work ability and increase the fairness and acceptance of the assessment process.

Competing interests

Simon Lauper is the director and Johanna Zwimpfer co-director of the Aerztliches Begutachtungsinstitut (ABI). Alain Nordmann has conducted paid expert disability assessments at the ABI. Tibor and Leon Zwimpfer are the children of the directors of the ABI and were employed to process and edit these data together with Benedikt Altermatt and Joerg Koehler. There are no other conflicts of interest to be declared.

Acknowledgments

The Basel Institute for Clinical Epidemiology and Biostatistics is supported by grants from Santésuisse and the Gottfried and Julia Bangerter-Rhyner-Foundation.

References

1. OECD Publishing. OECD Economic Outlook, Volume 2010 Issue 1. Chapter 5. Figure 5.19. Disability benefit recipient rates are high and still increasing in many countries. Version 1. Paris Cedex, France: Organization for Economic Co-operation and Development (OECD); 2010.
2. Federal Office for Social Insurance. Statistics of social security - disability insurance. Berne, Switzerland: Federal Office for Social Insurance; 2012. Available from: <http://www.bfs.admin.ch/bfs/portal/de/index/news/publikationen.html?publicationID=5204>.
3. Federal Office for Social Insurance. Background of multidisciplinary assessments in disability insurance. Berne, Switzerland: Federal Office for Social Insurance; 2012. Available from: <http://www.bsv.admin.ch/themen/iv/00027/index.html>.
4. Schellart AJ, Mulders H, Steenbeek R, Anema JR, Kroneman H, Besseling J. Inter-doctor variations in the assessment of functional incapacities by insurance physicians. *BMC Public Health*. 2011;11:864. <http://dx.doi.org/10.1186/1471-2458-11-864>.
5. Spanjer J, Krol B, Brouwer S, Groothoff JW. Sources of variation in work disability assessment. *Work*. 2010;37(4):405–11.
6. Marfeo EE, Eisen S, Ni P, Rasch EK, Rogers ES, Jette A. Do claimants over-report behavioral health dysfunction when filing for work disability benefits? *Work*. 2014 Mar 4. [Epub ahead of print].
7. Brouwer S, Dijkstra PU, Stewart RE, Goeken LN, Groothoff JW, Geertzen JH. Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. *Disabil Rehabil*. 2005;27(17):999–1005. <http://dx.doi.org/10.1080/09638280500052823>.
8. Dowrick AS, Gabbe BJ, Williamson OD, Wolfe R, Cameron PA. A comparison of self-reported and independently observed disability in an orthopedic trauma population. *J Trauma*. 2006;61(6):1447–52. <http://dx.doi.org/10.1097/01.ta.0000242219.02901.d9>.
9. Slebus FG, Kuijer PP, Willems JH, Frings-Dresen MH, Sluiter JK. Work ability assessment in prolonged depressive illness. *Occup Med (Lond)*. 2010;60(4):307–9. <http://dx.doi.org/10.1093/occmed/kqq079>.
10. Mueller JP, Reich J. Legal opinion of the compatibility of the Swiss Federal Court rulings on medical evaluation of claims for disability insurance benefits with Article 6 of the Convention for the Protection of Human Rights and Fundamental Freedoms (4 November 1950). Berne: Switzerland; 2010: <http://www.studerrechtsanwalt.ch/wp-content/uploads/2010/06/gutachten-2010-02-12-rechtsgutachten-dr-j-p-muller.pdf>.
11. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud*. 2011;48(6):661–71. <http://dx.doi.org/10.1016/j>

- ijnurstu.2011.01.016.
12. Uebersax J. Statistical Methods for Rater and Diagnostic Agreement. John Uebersax Enterprises LLC; 2010. <http://www.john-uebersax.com/stat/agree.htm>.
 13. Agresti A. An introduction to categorical data analysis. Second Edition. New York: John Wiley & Sons, Inc; 2002. p. 275–7. <http://dx.doi.org/10.1002/0471249688>.
 14. Zeger SL, Liang KY. Models for Longitudinal data: a Generalized Estimating Equation Approach *Biometrics*. 1988;44(1):1049–60.
 15. Papke LE WJ. Econometric Methods for Fractional Response Variables With an Application to 401 (K) Plan Participation Rates. *Journal of Applied Econometrics*. 1996;11(6):619–32. [http://dx.doi.org/10.1002/\(SICI\)1099-1255\(199611\)11:6<619::AID-JAE418>3.0.CO;2-1](http://dx.doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1).
 16. Federal Office for Social Insurance. Statistics of social security - Disability insurance. 2012. [p. 19]. Berne, Switzerland. Available from: <http://www.bfs.admin.ch/bfs/portal/de/index/news/publikationen.html?publicationID=5204>.
 17. Wind H, Samoocha D, van der Beek AJ, Frings-Dresen MH. Prevention of disability: The opinion of claimants applying for a disability benefit. *Work*. 2013 Jun 26. [Epub ahead of print].
 18. Winde LD, Alexanderson K, Carlsen B, Kjeldgard L, Wilteus AL, Gjesdal S. General practitioners' experiences with sickness certification: a comparison of survey data from Sweden and Norway. *BMC Fam Pract*. 2012;13:10. <http://dx.doi.org/10.1186/1471-2296-13-10>.
 19. Englund L, Svardsudd K. Sick-listing habits among general practitioners in a Swedish county. *Scand J Prim Health Care*. 2000;18(2):81–6. <http://dx.doi.org/10.1080/028134300750018954>.
 20. Stohr S, Bollag Y, Auerbach H, Eichler K, Imhof D, Fabbro T, et al. Quality assessment of a randomly selected sample of Swiss medical expertises--a pilot study. *Swiss Med Wkly*. 2011;141:w13173.
 21. Carey TS, Hadler NM, Gillings D, Stinnett S, Wallsten T. Medical disability assessment of the back pain patient for the Social Security Administration: the weighting of presenting clinical features. *J Clin Epidemiol*. 1988;41(7):691–7. [http://dx.doi.org/10.1016/0895-4356\(88\)90121-7](http://dx.doi.org/10.1016/0895-4356(88)90121-7).
 22. Slebus FG, Sluiter JK, Kuijer PP, Willems JH, Frings-Dresen MH. Work-ability evaluation: a piece of cake or a hard nut to crack? *Disabil Rehabil*. 2007;29(16):1295–300. <http://dx.doi.org/10.1080/09638280600976111>.
 23. Swiss Federal Court decision: BGE 130 V 352. (March 12, 2004). [in German]. Available from: <http://www.bger.ch/index/jurisdiction/jurisdiction-inherit-template/jurisdiction-recht-leitentscheide1954-direct.htm>.
 24. Pransky G, Katz JN, Benjamin K, Himmelstein J. Improving the physician role in evaluating work ability and managing disability: a survey of primary care practitioners. *Disabil Rehabil*. 2002;24(16):867–74. <http://dx.doi.org/10.1080/09638280210142176>.
 25. Letrilliart L, Barrau A. Difficulties with the sickness certification process in general practice and possible solutions: a systematic review. *Eur J Gen Pract*. 2012;18(4):219–28. <http://dx.doi.org/10.3109/13814788.2012.727795>.
 26. Anner J, Schwegler U, Kunz R, Trezzini B, de Boer W. Evaluation of work disability and the international classification of functioning, disability and health: what to expect and what not. *BMC Public Health*. 2012;12:470. <http://dx.doi.org/10.1186/1471-2458-12-470>.

Received for publication: 10 February 2014