Original article
Scand J Work Environ Health Online-first -article
doi:10.5271/sjweh.3440

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

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

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
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, MSc 1


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 Co-operation 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-
Assessing work ability of disability claimants

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: ≤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]

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

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).

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

---

**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].

<table>
<thead>
<tr>
<th></th>
<th>N (missing)</th>
<th>N</th>
<th>≤30%</th>
<th>31–50%</th>
<th>51–79%</th>
<th>80–99%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obs</td>
<td>Pred</td>
<td>Obs</td>
<td>Pred</td>
<td>Obs</td>
</tr>
<tr>
<td>Last job</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claimant</td>
<td>3462</td>
<td>100</td>
<td>83</td>
<td>83</td>
<td>10</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Physician</td>
<td>3286</td>
<td>276</td>
<td>80</td>
<td>81</td>
<td>15</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Expert</td>
<td>3556</td>
<td>6</td>
<td>45</td>
<td>40</td>
<td>18</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>Alternative job</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claimant</td>
<td>3354</td>
<td>208</td>
<td>74</td>
<td>74</td>
<td>17</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>Physician</td>
<td>2622</td>
<td>940</td>
<td>58</td>
<td>59</td>
<td>28</td>
<td>23</td>
<td>3</td>
</tr>
<tr>
<td>Expert</td>
<td>3558</td>
<td>4</td>
<td>7</td>
<td>12</td>
<td>17</td>
<td>19</td>
<td>14</td>
</tr>
</tbody>
</table>

**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.
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 47% 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 difference 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 (overlaying 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 reverse. Fine grey lines denote the disability categories used by the Swiss Accident Insurance Company: ≤30%=severe restriction, 31–50%=moderate-to-severe restriction, 51–79%=mild to moderate restriction, 80–99%=mild restriction, 100%=no restriction.
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 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]

<table>
<thead>
<tr>
<th>Comparison</th>
<th>OR ± 95% CI</th>
<th>MD % ± 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Last job</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician versus claimant</td>
<td>1.19 ± 1.09–1.30</td>
<td>1 ± 0.5–2.2</td>
</tr>
<tr>
<td>Expert versus claimant</td>
<td>7.38 ± 6.68–8.15</td>
<td>29 ± 28–31</td>
</tr>
<tr>
<td>Expert versus physician</td>
<td>6.20 ± 5.65–6.80</td>
<td>28 ± 26–29</td>
</tr>
<tr>
<td>Alternative job</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician versus claimant</td>
<td>2.01 ± 1.84–2.19</td>
<td>11 ± 10–12</td>
</tr>
<tr>
<td>Expert versus claimant</td>
<td>20.4 ± 16.5–22.4</td>
<td>57 ± 56–58</td>
</tr>
<tr>
<td>Expert versus physician</td>
<td>10.1 ± 9.29–11.1</td>
<td>40 ± 45–48</td>
</tr>
<tr>
<td>Alternative versus last job (patients)</td>
<td>1.70 ± 1.57–1.84</td>
<td>6 ± 5.1–6.9</td>
</tr>
<tr>
<td>Alternative versus last job (experts)</td>
<td>2.87 ± 2.64–3.11</td>
<td>16 ± 14–17</td>
</tr>
<tr>
<td>Alternative versus last job</td>
<td>4.70 ± 4.37–5.05</td>
<td>34 ± 33–35</td>
</tr>
</tbody>
</table>

*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.

---

**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].

<table>
<thead>
<tr>
<th>Last job</th>
<th>Without severe depression</th>
<th>With severe depression a</th>
<th>Difference between subgroups</th>
<th>Without somatoform disorder</th>
<th>With somatoform disorder b</th>
<th>Difference between subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MD % ± 95% CI</td>
<td>MD % ± 95% CI</td>
<td>MD % ± 95% CI</td>
<td>MD % ± 95% CI</td>
<td>MD % ± 95% CI</td>
<td>MD % ± 95% CI</td>
</tr>
<tr>
<td>Expert versus physician</td>
<td></td>
<td>25 ± 23–26</td>
<td>21 ± 17–24</td>
<td>-4 ± -7.6–-0.3</td>
<td>25 ± 23–26</td>
<td>34 ± 32–37</td>
</tr>
</tbody>
</table>

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

*Patients with somatoform disorder: 1123 (32% of 3561 without missing records - diagnoses with and without influence on work ability considered).
factors will influence work ability in both jobs. In inter
assessments by physicians between LJ and AJ, as these
experts. This could partly explain the similarity of
weight on environmental and personal factors than
experienced in disability assessments (21).

48 identical clinical vignettes were presented to 36 GP
substantial variability in disability assessments when
be of insufficient quality (20). In a US study, there was
of a random sample of 97 assessments were found to
ability, and liability insurance agencies in 2008, 20%
of 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). How-
ever, 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 inde-
pendent interdisciplinary expert team assess disability
claims to mitigate bias but multidisciplinary assessment
centers are paid by disability insurers and not by a neu-
tral third organization. We cannot rule out unconscious
bias by experts.

Fourth, expert assessment does not necessarily
ensure high quality and reproducible assessment of a claimant’s work ability. In a pilot study of expert
medical assessments submitted to Swiss accident, dis-
ability, 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 struc-
tures” 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 fac-
tors” (eg, workplace factors, conflicts with employer or
family) and “personal factors” (eg, coping, compliance
to therapy, illness behavior, motivation, age) were sel-
dom 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 excep-
tions to this rule where patients cannot be expected to
willfully 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 abil-
ity 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 codirector 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


Received for publication: 10 February 2014