To ask the right question at the right time:

When is the patient's self-assessed work ability most accurate as a predictor of the remaining duration of certified sickness absence?

Harald Reiso, MD[1,2], Jan F. Nygård, BSc, BA[3,4], Pål Gulbrandsen, MD, PhD[1], Sören Brage, MD, PhD[5] and Gunnar Tellnes, MD, PhD[1]

1) Department of General Practice and Community Medicine, Section for Occupational Health and Social Insurance Medicine, University of Oslo
2) Aust-Agder County Office of the National Insurance Service, Arendal, Norway
3) Department Group of Basic Medical Sciences, University of Oslo, Norway
4) The Cancer Registry of Norway
5) National Insurance Administration, Oslo, Norway

Correspondence: Harald Reiso, Aust-Agder County Office of the National Insurance Service, PO Box 1853 Stoa, N-4858 Arendal, Norway
Telephone: + 47 37 00 43 11 Telefax: + 47 37 00 43 01 E-mail: harald.reiso@samfunnsmed.uio.no

ABSTRACT

Aim: To present a model that estimates when predictors most accurately predict the remaining duration of certified sickness absence.
Methods: A questionnaire was given to 549 patients who were certified sick, and to their doctors. The questionnaires were completed at consultations that occurred from one to 20 weeks after the start of absence. The duration of the episodes of certified sickness absence was provided by the National Sickness Benefit Register. Four separate analyses were performed for self-assessed work ability, diagnoses, age and gender of the patients. The difference between observed remaining duration of the episodes of absence and their estimated expected duration, in percent, served as a measure of predictive accuracy. The improvement of predictive accuracy, compared with a variable that had no association with the outcome (a constructed comparison variable), was estimated for each of the variables of interest, by weeks of consultations.
Results: Self-assessed work ability had no improved predictive accuracy in consultations at one week. It was 10% better at eight weeks, reached a maximum of 12% at 13 weeks, and was 10% better at 20 weeks. The predictive accuracy of diagnoses was at its best of 22% in consultations at one week, and at a minimum of 9% better at 11 weeks. The predictive accuracy of age was 7% better at one week, and 10% at 20 weeks. Gender showed no improved predictive accuracy.
Conclusions: Self-assessed work ability is most accurate as a predictor in the interval between eight and twenty weeks from the start of absence, diagnoses are most accurate in the first weeks.

Key words: duration (of certified sickness absence), epidemiology, functional assessment, method, predictive accuracy, sick-leave, sick-listing, sickness absence, sickness certification, work ability

BACKGROUND

Medicine has to deal with new predictive factors, as advances in epidemiology, statistics and computing power have made increased use of prognostic models possible (1-3). Predictive factors have certain predictive magnitudes, measured by rate or hazard ratios. These ratios indicate whether values of independent variables are predictive, relative to reference categories, or not. Analyses may be performed at different times during e.g. a disease, or an episode of certified sickness absence (4,5). Such an approach does not give measures about how useful a prognostic factor per se is, that is how accurately the factor predicts for example duration of absence, nor how that accuracy changes over time. A factor with a low or unknown predictive accuracy is of limited use as a prognostic tool, and even strong predictors may have low predictive accuracy (3).

How the predictive accuracy of factors change over time may provide important information for deciding which interventions to choose, and at what time to use them. For example, when do measurements of the functional consequences of a disease most accurately serve as a predictor of the remaining duration of certified sickness absence? This information could be useful for doctors to initiate therapy and rehabilitation at the right time (6). Analyses that estimate the predictive accuracy of variables, and how it changes during the course of a disease, may thus supply useful prog-

nostic information.

We have shown in a previous study that self-assessed work ability predicts the remaining duration of certified sickness absence in episodes that have lasted more than one week at the time the assessments are made, in a representative group of patients in general practice (7). Diagnostic groups predict remaining duration at the start of episodes, but not later, while the age of patients predicts remaining duration both in new and prolonged episodes. The gender of the patients is not predictive of duration. However, although self-assessed work ability, diagnoses and the age of patients are significant predictors of duration of absence, their change of predictive accuracy by time from the start of sickness absence is not known. And thus it is not certain when assessments of these predictors should be made.

AIM

The aim of this study was to present a model that estimates when predictors most accurately predict the remaining duration of certified sickness absence.

METHODS

Data collection

Data collection took place January-April 1996 in the county of Aust-Agder, southern Norway (7). The county had 100,211 inhabitants in 1996. All general practitioners and company doctors in the county were invited to participate in the study. Each doctor recorded episodes of certified sickness absence made during office hours. Patients who were issued new sickness certificates, or certificates for prolongation of ongoing episodes of absence, were consecutively included in the study. No instruction was given that a certain time of absence episodes should be given special attention. Patients and doctors completed questionnaires independent of one another, and were assured that participation in the study would not interfere with absence status. For episodes recorded more than once, only the first was included in the analyses.

This study focused on prediction of duration early in the course of absence. Episodes with duration of more than 20 weeks at the time of consultations were therefore excluded.

The sample

From a sample of 567 paired questionnaires from patients and doctors, 11 were excluded because they concerned patients certified sick while on rehabilitation benefits, and seven were excluded for other reasons (7). The study sample consisted of questionnaires from 549 patients, 49 general practitioners, and three company doctors. The doctors recorded an average of 11 episodes each (minimum 4, maximum 23). The patients' ages ranged from 17-65 years (mean 41 years), and 56% were women.

Definitions

- *Accuracy* is defined as the degree to which a measurement, or an estimate based on measurements, represents the true value of the attribute that is being measured (8).
- *Predictive accuracy* in this paper is defined as the difference in percent between model predictions of duration (expected remaining duration) and observed remaining duration of episodes of certified sickness absence.
- *Improved predictive accuracy* is defined as the improvement of predictive accuracy for the variable of interest (estimated in percent), compared with a variable that had no association with the outcome (a constructed comparison variable).

Variables

The patients answered the following question about work ability: "To what degree is your ability to perform your ordinary, remunerative work reduced today?" The answer categories were: "Very much reduced", "much reduced", "moderately reduced", "not much reduced" and "hardly reduced at all" (9). As the distribution of these categories was strongly skewed, work ability assessed as moderately, not much, or hardly reduced at all was recoded to moderately reduced. The other options were left unchanged.

The doctors classified the main sickness certification diagnoses according to ICPC (International Classification of Primary Care) (10). Musculoskeletal, psychological, respiratory disorders, and injuries are important groups of diagnoses in sickness absence (11, 12). The diagnoses were therefore recoded into those groups, and all other disorders. The age of patients was recoded into four groups; 17-30 years, 31-40, 41-50 and 51-65 years.

An episode of certified sickness absence ends by return to work, by reaching maximum sickness benefit time – which is 365 calendar days in Norway, or by transition to other benefits. The observed remaining duration of absence, from the time of the consultation to the end of the episode, was measured in calendar days. The National Sickness Benefit Register supplied duration data (13). No episode was lost to follow-up.

Analyses

For details, see Appendix. In short, linear regressions, stratified by weeks of consultation (one to 20 weeks), were used to estimate the expected remaining duration for each episode of certified sickness absence, based on observed remaining duration (Stage I analyses). Self-assessed work ability, diagnoses, age and gender of the patients were independent variables in four separate analyses.

The difference in percent between expected and observed remaining duration served as a measure of the accuracy of the prediction. These accuracy measures were used in Stage II linear regressions with

weeks of consultation. Because the possible range of observed remaining duration *after* consultation is associated with the duration of certified sickness absence *before* consultation, predictive accuracy is time-dependent, which was corrected for by including second-order coefficients (see Appendix for details). Furthermore, estimations of how much better the predictive accuracy for the variables of interest were, compared with a variable that had no association with the outcome (a constructed comparison variable), were performed (see Appendix for details). The *improved* predictive accuracy was given in percent.

RESULTS

Work ability was assessed by the patients as moderately reduced in 29% of the cases, much reduced in 34%, and very much reduced in 37% (Table 1). The diagnostic groups were musculoskeletal disorders in 42% of the cases, psychological disorders in 18%, respiratory disorders in 16%, injuries in 5%, and all other disorders in 19%.

Self-assessed work ability had no improved predictive accuracy in consultations at one week. It was 10% better at eight weeks, reached a maximum of 12% at 13 weeks, and was 10% better at 20 weeks (Figure 1).

The predictive accuracy of diagnoses was at its best of 22% in consultations at one week, and at a minimum of 9% better at 11 weeks. The predictive accuracy of age was 7% better at one week, and 10% at 20 weeks. Gender showed no improved predictive accuracy.

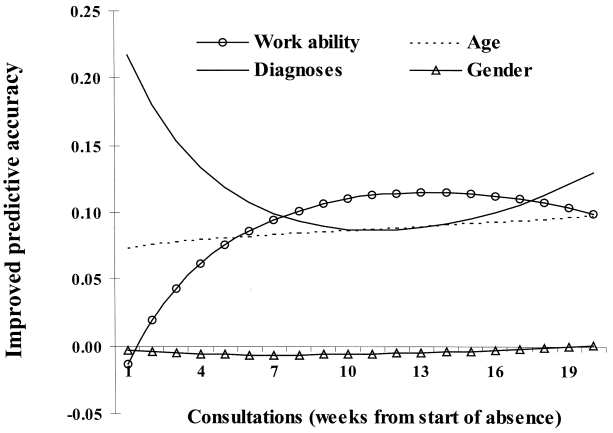

Figure 1. The improved predictive accuracy of self-assessed work ability, diagnoses, age and gender of the patients, by time of consultations (weeks). Improved predictive accuracy is the improvement of predictive accuracy for the variable of interest (estimated in percent), compared with a variable that had no association with the outcome (a constructed comparison variable). N = 549. Aust-Agder County, Norway, 1996.

Table 1. The distribution of the patients' self-assessed work ability, diagnoses, age and gender of the patients, and the mean observed remaining duration of the episodes of certified sickness absence (with 95% confidence intervals - 95% CI), by weeks of consultations. N = 549 (Aust-Agder County, Norway, 1996).

Consultations (weeks from start of absence)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
Work ability																					(Total)	(%)
Moderately reduced	92	14	12	9	–	3	–	3	2	1	2	3	1	5	1	1	1	1	2	4	(157)	(28.6)
Much reduced	93	11	19	6	4	3	3	5	6	5	5	2	6	3	3	6	3	2	3	1	(189)	(34.4)
Very much reduced	92	14	13	13	9	5	5	7	2	6	3	7	5	3	2	1	5	6	4	1	(203)	(37.0)
																					(549)	(100.0)
Diagnoses																						
Musculoskeletal disorders	89	20	23	12	7	7	5	11	5	6	4	5	9	5	4	4	7	1	4	3	(231)	(42.1)
Psychological disorders	38	6	10	5	2	1	2	1	3	4	3	1	2	5	1	2	1	6	2	2	(97)	(17.7)
Respiratory disorders	74	4	1	2	–	–	1	–	–	1	1	2	–	–	1	–	–	1	–	–	(88)	(16.0)
Injuries	19	1	1	3	1	1	–	1	1	–	–	–	1	–	–	–	–	–	1	–	(30)	(5.5)
All other disorders	57	8	9	6	3	2	–	2	1	1	2	4	–	1	–	2	1	1	2	1	(103)	(18.7)
																					(549)	(100.0)
Age																						
17-30 years	74	7	6	9	3	4	–	2	1	5	1	3	2	–	–	2	2	1	2	1	(125)	(22.8)
31-40 years	76	13	12	3	4	2	4	4	3	3	2	2	2	4	1	2	–	2	–	2	(141)	(25.7)
41-50 years	72	16	11	7	6	3	3	5	3	–	3	2	4	4	2	3	4	4	3	2	(157)	(28.6)
51-65 years	55	3	15	9	–	2	1	4	3	4	4	5	4	3	3	1	3	2	4	1	(126)	(22.9)
																					(549)	(100.0)
Gender																						
Male	126	14	19	14	5	10	6	8	5	4	3	7	4	2	1	5	1	4	4	1	(243)	(44.3)
Female	151	25	25	14	8	1	2	7	5	8	7	5	8	9	5	3	8	5	5	5	(306)	(55.7)
Total number of observations	277	39	44	28	13	11	8	15	10	12	10	12	12	11	6	8	9	9	9	6	(549)	(100.0)
Mean observed remaining duration (calendar days)	43	67	78	81	91	168	203	156	157	148	94	168	157	147	192	150	198	186	164	150		
(95% CI lower)	34	33	49	47	54	101	115	89	95	76	52	107	100	86	113	86	147	126	114	88		
(95% CI upper)	53	102	107	114	127	235	291	224	218	221	137	229	213	208	270	214	250	246	215	212		

DISCUSSION

Method

Participating patients were similar to other patients certified sick regarding age, gender and diagnoses (7,9). Responding patients were similar to non-responding patients, as were participating to non-participating doctors. The National Sickness Benefit Register in Norway serves as an account system for the payment of sickness benefits, and is therefore revised, audited and quality controlled. The register also ensures good follow-up of the cases. None were lost to follow-up in this study.

There were few observations in consultations at more than four weeks (Table 1). This should be considered when interpreting the results, because an estimation based on one observation yields high predictive accuracy, i.e. estimated equals observed remaining duration.

The presented study resembles meta-analysis, because it can be considered as 20 different studies, one for each week, and the regression a meta-analysis that estimates an overall predictive accuracy for the variable of interest. A test for homogeneity would check for differences of accuracy between the studies, but only yield a dichotomous answer (14). Cox models with time-dependent co-variates, and non-parametric regression spline models, can take changes of hazard that may occur after the assessments of the predictors into account, by using repeated measurements of the predictor (3, 4). Such approaches applied to our data would neither measure how accurately the variable of interest predicts duration, nor how that accuracy changes by the time of consultation, and would represent even more complicated analyses than the ones presented here. Our approach estimates the change in association over time.

Main findings

When trying to assess the prognosis of a disease it is necessary to take the time of the assessment, relative to the start of the disease, into account. Prognostic tools should be predictive and accurate. We consider an improved predictive accuracy of self-assessed work ability of 12%, 13 weeks after the start of certified sickness absence, to be of possible importance in clinical practice. Diagnoses had an improved predictive accuracy at the beginning of absence, age had a gradual increase of accuracy by time of consultations, while gender was neutral. Thus, different information has different prognostic value by time.

Other possible applications

Methods estimating predictive accuracy may enable doctors to make more confident statements about prognoses, based on measurements performed at different times during the course of diseases. The presented model may thus be useful in other fields of medicine that deal with time-dependent data. For example, when during hepatitis do the levels of liver enzymes most accurately predict the duration of the disease?

CONCLUSIONS

Self-assessed work ability is most accurate as a predictor of the remaining duration in the interval between eight and twenty weeks from the start of absence, diagnoses are most accurate in the first weeks.

ACKNOWLEDGEMENTS

The authors thank Håkon K. Gjessing for useful discussions. The Norwegian Ministry of Health and Social Affairs supported the study financially. The study was approved by the Regional Ethics Committee for Medical Research, the Norwegian Data Inspectorate, and the Legal Affairs Division of the National Insurance Administration.

APPENDIX

Stage I analyses

Linear regressions were used to estimate the expected remaining duration ($D_{expected}$) for each episode of absence, for the variables of interest, by when consultations occurred (T_0=1,2,…20 weeks), based on the observed remaining duration ($D_{observed}$) of the episode. 20 separate, ordinary linear regression analyses were performed with self-assessed work ability, diagnoses and age and gender of the patients as independent variables in four sets of analyses.

The ratio of expected remaining duration to observed remaining duration ($D_{expected} / D_{observed}$), served as a measure of the accuracy of the prediction. Because a regression yields the mean of the dependent variable as the optimal expected value, approximately half of the observations will be longer, and half will be shorter than the expected value. The direction of such a deviation was not our interest. We wanted a measure of deviation from perfect accuracy. Values of $D_{expected} / D_{observed}$ of less than 1.0 were kept unchanged, while values larger than 1.0 were inverted. The portions were multiplied by 100, which gives the deviation from perfect accuracy in percent.

Stage II analyses

Correction for time dependency

Because there is a maximum possible duration of certified sickness absence of 365 days, there is a non-linear increase in accuracy depending on the time of consultation. At consultation of one week (T_0 = 1), observed remaining duration can vary between one and 52 weeks, while at 20 weeks it can vary between one and 32 weeks. Observed remaining duration thus depends on when the consultation occurs, as do expected remaining duration. The probability that expected remaining duration equals observed remaining duration is less likely by coincidence at consultation at one week than at 20 weeks, and the predictive accuracy therefore also depends on when the consultation occur. The difference between predictive accuracy for a variable of interest in consultations at 19 and 20 weeks is larger than the difference at one and two weeks. Therefore the change in predictive accuracy, by when consultations occur, is non-linear. The simplest non-linear model, using a second order factor, was introduced in the Stage II analyses to correct for this.

Linear regressions with second order factors were fit to the Stage I accuracy measures and time of consultations (one to 20 weeks) for the variables of interest (Stage II analyses). The transformed measures of $D_{expected}$ / $D_{observed}$ was the dependent variable, T_0 with second order factor was the independent variable. The resulting second (β_2) and first (β_1) order coefficients, and the constant term (k), are presented in Table 2.

Table 2. The results of ordinary linear regression analyses, with predictive accuracy in consultations at one to 20 weeks as dependent variable (see Equation 1 and 2), for the variables of interest and the comparison variable (coefficients are multiplied by 100 to ease their readability). Aust-Agder County, Norway, 1996.

	n	β_2	95% CI of β_2	β_1	95% CI of β_1	k	95% CI of k
Work ability assessed by patients	549	-0.14	-0.22 – -0.05	5.18	3.73–6.62	23.4	19.9–26.9
Diagnostic groups	549	-0.07	-0.16 – 0.02	3.60	2.00–5.19	31.6	27.7–35.4
Age-groups of patients	549	-0.11	-0.19 – -0.02	4.44	2.92–5.95	26.6	23.0–30.3
Gender of patients	549	-0.10	-0.18 – -0.01	3.99	2.53–5.46	24.9	21.3–28.4
Comparison variable		-0.10		4.06		24.9	

These β_2s, β_1s, and ks were used to calculate the predictive accuracy for the variable of interest (X), by weeks of consultation (T_0), as noted in Equation 1.

$$Accuracy_X(T_0) = \beta_{2,X}T_0^2 + \beta_{1,X}T_0 + k_X \quad (T_0=1,2,\ldots,20) \qquad \text{Equation 1}$$

Estimation of improved predictive accuracy

In order to yield a measure of how much better the predictive accuracy for the variable of interest was, compared with a variable that had no association with the outcome, a comparison variable was constructed. Variables with random values with a normal distribution, a mean of zero and a standard deviation of one, and with a uniform distribution between zero and one, were used in that construction. The means of 1000 Stage I analyses, by weeks of consultation (one to 20), using 1000 different generated variables with random values, were used in Stage II analyses. The means represent the most probable values for the predictive accuracy of any variable, by when the consultations occurred. The use of normally or uniformly distributed random values gave similar means. The uniform distribution was arbitrarily chosen. The resulting β_2, β_1 and k are presented in Table 2. The estimation of the comparison variable (Z), is expressed in Equation 2, where l is the number of regressions, and j each episode at the time of consultations.

$$\hat{D}_{\exp ected_{T_0,j}}(Z) = \left(\frac{\sum_{l=1}^{l=1000}\hat{\beta}_{2,T_0,j}}{1000}\right)Z^2 + \left(\frac{\sum_{l=1}^{l=1000}\hat{\beta}_{1,T_0,j}}{1000}\right)Z + \left(\frac{\sum_{l=1}^{l=1000}\hat{k}_{T_0,j}}{1000}\right) \qquad \text{Equation 2}$$

The predictive accuracy for the variable of interest (X) was compared with the comparison variable (Z), yielding improved accuracy, as noted in Equation 3.

$$\text{Improved accuracy } (T_0) = \frac{\beta_{2,X}T_0^2 + \beta_{1,X}T_0 + k_X}{\beta_{2,Z}T_0^2 + \beta_{1,Z}T_0 + k_Z} \quad (T_0=1,2,\ldots,20) \qquad \text{Equation 3}$$

This measure was subtracted 1.00, which means that equal values of the ratio components yielded zero. The deviation from zero was plotted against when the consultations occurred (Figure 1), which visualized the improvement in percent in predictive accuracy for the variables of interest.

REFERENCES

1. Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000; **29**: 158-67.
2. Aalen OO. Medical statistics – no time for complacency. *Stat Methods Med Res* 2000; **9**: 31-40.
3. Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. *Biometrics* 2000; **56**: 249-55.
4. Heinzl H, Kaider A, Zlabinger G. Assessing interactions of binary time-dependent covariates with time in Cox proportional hazards regression models using cubic spline functions. *Stat Med* 1996; **15**: 2589-601.
5. Quantin C, Abrahamowicz M, Moreau T, Bartlett G, MacKenzie T, Tazi MA, et al. Variation over time of the effects of prognostic factors in a population-based study of colon cancer: Comparison of statistical models. *Am J Epidemiol* 1999; **150**: 1188-200.
6. http://www.dep.no/sos/norsk/publ/utredninger/NOU/ Governmental report NOU 2000: 27, *Sickness absence and disability pension, an including working-life* (in Norwegian) (13.02.2003).
7. Reiso H, Nygård JF, Brage S, Gulbrandsen P, Tellnes G. Work ability and duration of certified sickness absence. *Scand J Public Health* 2001; **29**: 218-25.
8. Last JM. *A Dictionary of Epidemiology*. Third edition. New York: Oxford University Press, 1995.
9. Reiso H, Nygård JF, Brage S, Gulbrandsen P, Tellnes G. Work ability assessed by patients and their GPs in new episodes of sickness certification. *Fam Pract* 2000; **17**: 139-44.
10. Brage S, Bentsen BG, Bjerkedal T, Nygård JF, Tellnes G. ICPC as a standard classification in Norway. *Fam Pract* 1996; **13**: 391-6.
11. Söderberg E, Alexanderson K. *Literature study – published studies of the interface between medical practice and assessments of social insurance legislation*. Försäkringsmedicinskt centrum, Rapport 2001: 1. Linköping, 2001 (in Swedish).
12. Tellnes G, Svendsen KOB, Bruusgaard D, Bjerkedal T. Incidence of sickness certification. Proposal for use as a health status indicator. *Scand J Primary Health Care* 1989; **7**: 111-7.
13. Brage S, Nygård JF, Tellnes G. The gender gap in musculoskeletal-related long term sickness absence in Norway. *Scand J Soc Med* 1998; **26**: 34-43.
14. Fleiss JL. The statistical basis for meta-analysis. *Stat Methods Med Res* 1993; **2**: 121-45.