Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Health Economics 22 (2003) 747–762
JOURNAL OF HEALTH ECONOMICS
www.elsevier.com/locate/econbase

A low-key social insurance reform—effects of multidisciplinary outpatient treatment for back pain patients in Norway

Arild Aakvik[a,c,*], Tor Helge Holmås[a,c], Egil Kjerstad[b,c]

[a] *Department of Economics, University of Bergen, Fosswinckelsg 6, N-5007 Bergen, Norway*
[b] *Institute for Research in Economics and Business Administration (SNF), Breiviksveien 40, N-5045 Bergen, Norway*
[c] *Programme for Health Economics in Bergen (HEB), Department of Economics, Fosswinckelsg 6, N-5007 Bergen, Norway*

Received 22 March 2002; received in revised form 24 October 2002; accepted 6 February 2003

Abstract

This paper estimates treatment effects for back pain patients using observational data from a low-key social insurance reform in Norway. Using a latent variable model, we estimate the average treatment effect (ATE), the average effect of treatment on the treated (TT), and the distribution of treatment effects for multidisciplinary outpatient treatment at three different locations. To estimate these treatment effects, we use a discrete-choice model with unobservables generated by a factor structure model. Distance to the nearest hospital (in kilometres) is used as an instrument in estimating the different treatment effects. We find a positive effect of treatment of around 6 percentage points on the probability of leaving the sickness benefits scheme after allowing for selection effects and full heterogeneity in treatment effects. We also find that there are sound arguments for expanding the multidisciplinary outpatient programme for treating back pain patients.

JEL classification: C35; H43; I10; J64

Keywords: Heterogeneous treatment effects; Discrete-choice models; Back pain patients; Multidisciplinary treatment

1. Introduction

In many countries, the social insurance system is under pressure from an aging population and an increased number of people on disability pensions. Absence through sickness

* Corresponding author. Tel.: +47-555-89200; fax: +47-555-89210.
E-mail address: arild.aakvik@econ.uib.no (A. Aakvik).

0167-6296/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0167-6296(03)00025-0

represents the third major type of financial transfer from the social insurance system to individuals. There is also a worrying connection between long-term absence through sickness and progression to a disability pension. Thus, reducing absence due to sickness is high on the political agenda since it will lessen the burden facing the social insurance system.

One of the main reasons for absence through sickness in Norway is back pain.[1] Traditionally, treatment of patients with back pain has not been prioritized in Norwegian hospitals, resulting in long waiting lists for inpatient treatment. As a consequence, people with back pain are on sick leave for relatively long periods of time. However, clinical evidence emerging in the last 5–10 years shows that multidisciplinary *outpatient* care (by medical doctors, physiotherapists or psychologists working in teams) gives promising results regarding function and pain, and the transition back to work for people with back pain.[2]

Partly on the basis of these results, the Norwegian government decided that one possible way forward in order to lessen the burden on the social insurance system is to boost the number of patients treated at outpatient clinics. To this end, the Norwegian government settled for financial incentives by introducing a new and higher outpatient tariff for multidisciplinary treatment of patients with back pain. The aim is two-fold: first, to create incentives for hospitals to establish designated outpatient clinics for back pain patients or to increase capacity in already established clinics; and second, to reward multidisciplinary outpatient treatment. The rationale for the reform is that increased capacity and increased utilization of treatment techniques with a high expected success rate should result in shorter sickness periods, and hence reduced sickness payments from the social insurance system.

In this paper, we analyse whether the low-key reform can be termed a success. We have taken the following approach: The main part of this paper deals with the estimation of multidisciplinary treatment effects for patients with back pain (mainly disk herniation and non-specific low back pain) using observational data from three different locations. We estimate an econometric model for evaluating treatment effects when outcomes are discrete and estimate a flexible model where responses to treatment vary among observationally identical persons. The outcome variable is a dichotomous variable indicating whether the patient leaves the sickness benefits scheme after 9 months. Our structural model can be used to generate a variety of mean treatment effects (the average treatment effect (ATE) and treatment effect on the treated (TT)) from a common set of parameters as well as distributions of treatment effects.

We address four questions in this paper: First, what types of patients are being treated in an outpatient hospital, of those who have been examined at an outpatient hospital? Second, what is the overall effect of treatment on the probability of leaving the sickness benefits scheme after 9 months? Third, which groups of individuals benefit most from treatment? Fourth, how important is it to control for observables and unobservables in understanding the selection and outcome processes? We also present a small-scale policy simulation of the effect of expanding the programme on the sickness benefits scheme.

[1] Hagen and Thune (1998) analyse work incapacity owing to low back pain in the general population using cases identified from files of the national medical insurance system. Their studies reaffirm the burden of low back pain to society.

[2] See Indahl et al. (1995), Haldorsen et al. (2002), and Skouen et al. (2002) for evaluations of back pain treatment in Norway. Guzmán et al. (2001), Karjalainen et al. (2001), and van Tulder et al. (2001) provide systematic literature reviews of randomized controlled trials related to the effect of multidisciplinary biopsychosocial rehabilitation.

The treatment effect, without adjusting for observed and unobserved selection into treatment, is 7.3 percentage points. Of those undergoing treatment, 48.1% left the sickness benefits scheme after 9 months, while 40.8% of those not under treatment left the sickness benefits scheme after 9 months. After allowing for observed characteristics, we find a treatment effect for those who are treated of 9.3 percentage points. However, after running a very flexible selection model that accounts for heterogeneous treatment effects, we find an effect of treatment on the treated of 6.3 percentage points.

In addition, we find that the average treatment effect, which is the effect for a randomly selected person from the pool of eligible patients, is higher than the TT. Adjusting for observed characteristics gives an average treatment effect of 12.3 percentage points, while adjusting for observed and unobserved selection gives an ATE of 9.5 percentage points. This indicates that expanding treatment slots may increase the overall benefits of treating back pain patients, since ATE is greater than TT.

This paper is organized as follows. In Section 2, we present a latent variable model that is related to the classical model of potential outcomes. The specification can be used to estimate structural econometric models. We define commonly used treatment effect parameters in terms of the latent variables. Section 3 presents background information on the treatment programme and the data used in the empirical section. In Section 4, we discuss the selection process for treatment using a probit model. In Section 5, we present the main estimation results from the model. Section 6 concludes the paper.

2. A latent variable model

For each person i we have two potential outcomes (Y_{0i}, Y_{1i}) corresponding to the potential outcomes in the untreated and treated states respectively. Let $D_i = 1$ denote the receipt of treatment and $D_i = 0$ denote non-receipt. Let Y_i be the measured outcome variable so that

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}.$$

This is the classical model of potential outcomes: see Neyman (1923), Fisher (1935), Roy (1951), Cox (1958), Quandt (1972), Rubin (1978) and Heckman and Honoré (1990). The model can be used to estimate structural econometric models. The model has two potential outcome states of which only one is observed for each individual.

We specify a discrete-choice framework where the unobserved heterogeneity is assumed to follow a factor structure. The decision rule for outpatient treatment is given by

$$D_i^* = Z_i \beta_D + U_{Di}, \quad D_i = \begin{cases} 1, & \text{if } D_i^* \geq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where D_i^* is a latent index that determines whether treatment is received or not, Z_i is a vector of background variables, β_D is a set of parameters that reflects the effect of changes in background variables on the treatment index, and U_D are the unobservables.

We specify an outcome equation that depends on whether the individual is in the treated or non-treated state. We have the following outcome equation for the treatment state

$$Y_{1i}^* = X_i\beta_1 + U_{1i}, \quad Y_{1i} = \begin{cases} 1, & \text{if } Y_{1i}^* \geq 0, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where Y_{1i}^* is the latent index of leaving the sickness benefits scheme after 9 months, and X_i is a vector of background variables that affect the outcome. X_i and Z_i are not necessarily the same vectors. In particular, we have included a variable in Z_i that is not included in X_i. The identifying exclusion restriction used is the distance in kilometres to the nearest hospital that treats back pain patients. This variable will be discussed in more detail in Section 5. The outcome in the non-treatment state is

$$Y_{0i}^* = X_i\beta_0 + U_{0i}, \quad Y_{0i} = \begin{cases} 1, & \text{if } Y_{0i}^* \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The effects of the unobservables are the same in both states if $U_{1i} = U_{0i}$. In this case, individuals with the same observed X will have the same treatment effect. Such a model is termed the common coefficient model: see Heckman (1978). In this paper, we assume $U_{1i} \neq U_{0i}$ and thus allow for an idiosyncratic gain from treatment for each individual. Thus, the model allows for treatment effects to vary by unobserved individual characteristics. This is a random coefficient model if patients act on U_{1i} and U_{0i}: see Heckman (1997). To build a structural random coefficient model, we assume the following structure for the error terms:

$$U_{Di} = \alpha_D\theta_i + \epsilon_{Di}, \tag{4}$$

$$U_{1i} = \alpha_1\theta_i + \epsilon_{1i}, \tag{5}$$

$$U_{0i} = \alpha_0\theta_i + \epsilon_{0i}, \tag{6}$$

where $\epsilon_D, \epsilon_1, \epsilon_0$, and θ have mean zero, are mutually independent, and are independent of the exogenous variables in the model: see Aakvik et al. (2000). The factor structure assumption for discrete-choice models was introduced by Heckman (1981). The parameter α_D is the factor loading for the selection outcome, α_1 the factor loading for the outcome equation with treatment, and α_0 the factor loading for the outcome equation in the non-treated state. The interpretation of this specification considers θ, which is common to all states, to be an unobserved covariate that affects the outcomes, and the α's to be regression coefficients. From the model, we can formulate several interesting treatment effect parameters within the framework of a flexible but parsimonious specification: see, for instance, Aakvik et al. (2000).

To identify the model, we assume $\alpha_D = 1$ and that θ follows the standard normal distribution. The standard normality assumption of θ is not needed: see Aakvik et al. (2000). We assume access to an i.i.d. sample and suppress the i subscript. We focus on three correlations, derived from Eqs. (4)–(6), for the unobservables in the model

$$\text{Corr}(U_0, U_1) = \frac{\text{Cov}(U_0, U_1)}{\sqrt{\text{Var}(U_0)}\sqrt{\text{Var}(U_1)}} = \frac{\alpha_1\alpha_0}{\sqrt{1+\alpha_0^2}\sqrt{1+\alpha_1^2}}, \tag{7}$$

$$\text{Corr}(U_D, U_0) = \frac{\text{Cov}(U_D, U_0)}{\sqrt{\text{Var}(U_D)}\sqrt{\text{Var}(U_0)}} = \frac{\alpha_0}{\sqrt{2}\sqrt{1+\alpha_0^2}}, \tag{8}$$

$$\text{Corr}(U_D, U_1) = \frac{\text{Cov}(U_D, U_1)}{\sqrt{\text{Var}(U_D)}\sqrt{\text{Var}(U_1)}} = \frac{\alpha_1}{\sqrt{2}\sqrt{1+\alpha_1^2}}, \tag{9}$$

which are easy to verify given our assumptions. We also have

$$\text{Cov}(U_0, \theta) = \alpha_0 \qquad \text{Cov}(U_1, \theta) = \alpha_1 \qquad \text{Cov}(U_D, \theta) = \alpha_D, \tag{10}$$

since $\text{Var}(\theta) = 1$.

In what follows, we approximate the distribution of θ with a finite number of support points. This is a common estimation strategy: see Butler and Moffitt (1982).

In a three-equation model with dichotomous outcomes, we can form the following equations. First, the probability of leaving the sickness benefits scheme in the treated state is given by

$$\Pr(Y_1 = 1|X) = \sum_{j=1}^{m} \pi_j \Phi(X\beta_1 + \alpha_1\theta_j), \tag{11}$$

where m is the number of support points, π the mass probabilities, which sum to 1, and Φ the standard normal cumulative distribution function. The probability of leaving the sickness benefits scheme in the non-treated state is given by

$$\Pr(Y_0 = 1|X) = \sum_{j=1}^{m} \pi_j \Phi(X\beta_0 + \alpha_0\theta_j). \tag{12}$$

This set-up is very flexible, since we allow $\beta_1 \neq \beta_0$ and $\alpha_1 \neq \alpha_0$. The probability of being treated in an outpatient hospital is given by

$$\Pr(D = 1|Z) = \sum_{j=1}^{m} \pi_j \Phi(Z\beta_D + \alpha_D\theta_j). \tag{13}$$

Eqs. (11)–(13) is a structural model in the sense that we can predict the outcome in the treated and non-treated state for each individual even if we do not observe each individual in both states. To define the effect of treatment on the treated we have to condition on $D = 1$. Using Bayes' rule, we get

$$\Pr(Y_1 = 1|D = 1, X) = \frac{1}{\Phi(Z\beta_D)} \sum_{j=1}^{m} \pi_j \Phi(X\beta_1 + \alpha_1\theta_j)\Phi(Z\beta_D + \alpha_D\theta_j), \tag{14}$$

and

$$\Pr(Y_0 = 1|D = 1, X) = \frac{1}{\Phi(Z\beta_D)} \sum_{j=1}^{m} \pi_j \Phi(X\beta_0 + \alpha_0\theta_j)\Phi(Z\beta_D + \alpha_D\theta_j). \tag{15}$$

The average treatment effect and the effect of treatment on the treated is given by

$$\text{ATE}(X) = \Pr(Y_1 = 1|X) - \Pr(Y_0 = 1|X), \tag{16}$$

and

$$\text{TT}(X) = \Pr(Y_1 = 1|D = 1, X) - \Pr(Y_0 = 1|D = 1, X), \tag{17}$$

see Heckman and Robb (1985) and Heckman (1997). Heckman and Vytlacil (2000) discuss the relationship between different treatment parameters. To find the average treatment effect we insert Eqs. (11) and (12) into Eq. (16), and to find the effect of treatment on the treated, we insert Eqs. (14) and (15) into Eq. (17). To find ATE, we average ATE(X) over the full sample. To find TT we average TT(X) over the sample of treated patients ($D = 1$).

An alternative approach to the latent variable framework would be to focus on the expected duration of the sickness benefits periods—for example, estimating the treatment effects in a duration model. However, since our data are non-experimental, we have to model the selection process for treatment. This will result in a very complicated likelihood function: see Ham and LaLonde (1996). One advantage of using a duration model approach is the possibility of testing for the duration of the treatment effect. This is difficult in our case, since we are not able to follow the individuals over a sufficiently long time period. Due to this limitation and to the modelling complexities, we have chosen to use the discrete-choice model. An advantage of our approach is that our model allows the treatment effect to vary with both observed and unobserved individual characteristics. The resulting estimates of the treatment effects in this model are easily compared with results from studies of randomized clinical trials for back pain patients.

Observational studies analyse the effect of treatment in a real-world situation. The advantage of this approach compared to randomized clinical trials is that we can address interesting policy questions other than the mean effect of treatment on medical outcomes estimated in randomized controlled trials. For instance, knowing what the actual selection process looks like may be interesting: who receives treatment and what are the important differences in individual characteristics between treated and non-treated patients. Observational studies can also be used to compare treatment effects under real-world conditions with treatment effects under ideal experimental conditions. Observational studies often use register data, which are less costly to collect than running a controlled experiment, and are often more suited to analysing economic questions such as the return to work. Controlled experiments face similar selection problems as non-experimental studies, if one analyses outcomes that are conditional on health outcomes: see Ham and LaLonde (1996).

3. Institutional settings and data

In Norway the sickness benefits system is organized under the National Insurance Administration (NIA). Sickness insurance is mandatory, covering all employees who have been with the same employer for at least 2 weeks. For sick leave lasting more than 3 days, a medical certificate is required. After 8 weeks, the employee is obliged to provide a more detailed certificate to the NIA, stating the diagnosis and a prognosis assessment. Sickness benefits are given for a maximum period of 1 year, the first 16 days being paid by the employer and the remaining period by the NIA. Coverage is 100% from the first day, but with an income ceiling at NOK 294,540 in 2000. However, since most large firms and the public sector compensate income above the insurance ceiling, the 100% replacement rate applies to the majority of employees. The Norwegian sickness benefits scheme is generous compared to schemes in other countries: see NOSOSCO (2001). Individuals unable to return to work after 1 year may apply for a disability pension or temporary rehabilitation benefits. These

benefits are considerably lower than sickness benefits, and the size of the benefit depends on previous work experience and income.

For several years, clinical research has aimed to establish new cause-and-effect mechanisms behind back pain illnesses and to evaluate the effect of different treatment programmes. Guzmán et al. (2001) find strong evidence that intensive multidisciplinary biopsychosocial rehabilitation of patients with chronic low back pain improves function when compared with inpatient or outpatient non-multidisciplinary treatment. They find some evidence that this type of rehabilitation reduces pain when compared with outpatient non-multidisciplinary treatment. With particular relevance to our study, they find contradictory evidence regarding vocational outcomes of intensive multidisciplinary biopsychosocial rehabilitation. Some trials reported improvements in work readiness, but others showed no significant reduction in sick leave. Karjalainen et al. (2001) find that there is some evidence showing that multidisciplinary rehabilitation for sub-acute low back pain, including work-site visits, is effective. The authors conclude that more studies on biopsychosocial multidisciplinary rehabilitation for sub-acute low back pain are needed, though. van Tulder et al. (2001) discuss whether behavioural therapy is more effective than reference treatments for chronic (disabling) low back pain. They conclude that behavioural treatment seems to be an effective treatment for patients with chronic low back pain. However, they find that it is not known what type of patients benefit most from different types of behavioural treatment.

Clinical researchers have succeeded in their aim of establishing new cause-and-effect mechanisms and to evaluate the effect of different treatment programmes in the sense that advice given by the medical professionals to people with back pain is slowly changing. From a period dominated by what can be termed passive treatment and a belief that rest and minimal physical activity would eventually free the patient from pain, the current strategy is early intervention and light physical activity.[3] Advice and instruction concerning how to cope with different diagnoses is an important part of treatment programmes. Patients are given information on the reason for their pain and why it hurts, and thereby also motivating them to undertake light exercise even if the pain is relatively severe. Most back pain-related illnesses do in fact disappear in a relatively short period of time and surgical interventions should in many cases be avoided. The best treatment is to motivate the patients to exercise and to reduce their anxiety by providing information.

In Norway, multidisciplinary biopsychosocial rehabilitation is provided at outpatient clinics. Examples of interventions are exercises, psychosocial training, and psychological counselling. The Norwegian government has established a so-called back pain network, the main activities of which are research into, and dissemination, of the effects of multidisciplinary biopsychosocial rehabilitation. The network provides clinical practice guidelines on multidisciplinary rehabilitation to general practitioners and specialists. The guidelines are based on the Royal College of General Practitioners' guidelines in the UK, which was

[3] The timing of interventions depends on the type of diagnosis made. One important distinction is between acute low back pain in the absence of "red flags" (e.g. signs of infection or cancer) and sub-acute and chronic low back pain. Acute low back pain, without "red flags", is recognized with periods of pain of high intensity followed by periods that are less painful. In the Norwegian clinical practice guidelines, referral to outpatient clinic is recommended if the patient is still on sick leave after 8–12 weeks. Manipulation may be adequate for some patient even earlier, perhaps only after 1–2 weeks.

originally based on the 1994 guidelines provided by the Agency for Healthcare Research and Quality in the US.

Treatment programmes based on clinical practice guidelines on multidisciplinary rehabilitation are what the Norwegian government wants to encourage by the new outpatient tariff. By differentiating the tariff, so that clinics get paid twice as much for treating patients on sick leave compared to other clients, the government is signalling that it wants clinics to be selective regarding the patient on whom their scarce resources are used.

In this study, we use a data set drawn from the Norwegian Patient Register (NPR). NPR is a large database containing patient data from all public general hospitals in Norway, as well as from some private clinics. The register provides detailed information on variables such as age, gender, medical diagnosis,[4] treatments, and the date of hospitalization. Since the register does not contain social security information, we have merged the data from the NPR with data from the NIA. Along with other information, these data provide us with exact dates for each patient's sick leave between 1 January 2000 and 31 December 2000.

Four conditions must be met for a patient to be part of the sample. First, the patient was examined (but not necessarily treated) at an outpatient spinal clinic in 2000. Second, the patient had a sub-type of the M-diagnosis (musculoskeletal pain) according to the ICD-10 classification. We used medical expertise to select the right medical diagnosis from NPR. Third, the patient was eligible for sickness benefits from the Norwegian mandatory sickness insurance system, and commenced a period of sick leave during the first 3 months of 2000. Fourth, the patient had not been on sick leave in the 6-month period prior to entering the sick leave period in 2000.

These criteria generate a sample of 656 individuals. Since the sampling period is the first 3 months of 2000, and since we have social security data until 31 December 2000, we can track the individuals in our sample for a period of at least 9 months. Obviously, there is a trade-off between the length of the sampling period and the follow-up period. A longer sampling period gives a larger sample but a shorter observation period. We have experimented with different lengths of the sampling period up to 6 months. The results are relatively robust to different combinations of the sampling period and follow-up period. Results from this type of sensitivity analysis are available from the authors.

In Table 1, we define the variables used in this paper. The outcome variable (sickness) is a dummy variable, which takes the value 1 if the individual leaves the sickness benefits scheme after 9 months, and 0 otherwise. Whether patients receive treatment or not is also represented by a dummy variable. The treatment variable equals 1 if the patient receives one or more treatments, and 0 otherwise. Further, we have information on age and gender, whith age being a continuous variable. We also have detailed information on medical diagnoses. The largest and most important groups of medical diagnoses are disk herniation[5] and low

[4] However, we are not able to distinguish between patients according to whether the reason for their sick leave is acute, sub-acute or chronic low back pain.

[5] A disk may herniate because of sudden trauma—anything from a fall on an icy sidewalk to an athletic injury to simply lifting the wrong bag of groceries in the wrong way at the wrong time. It may also be caused by the cumulative long-term effects of what doctors like to call poor body mechanics—a lifetime of too much bending and twisting in too many awkward positions. Disks herniate most commonly in the lower back, although this also occurs frequently in the lower neck and, less commonly, in other places.

Table 1
Variables used in the regressions

Variable name	Definitions
Treatment	Dummy variable for treatment (1 = treated, 0 = untreated)
Out of sickness	Dummy variable for leaving sickness benefits (1 = out of sickness, 0 = not out of sickness)
Distance	Distance in kilometres (in logarithms) to nearest hospital treating patients
Male	Dummy variable (1 = male, 0 = female)
Age	Age in years
Age^2	Age squared
Disk herniation	Medical diagnoses (1 = disk herniation, 0 = otherwise)
Low back pain	Medical diagnoses (1 = low back pain, 0 = otherwise)
Income 1999	Income in 1999, NOK 1000
Sick days 1999	Number of sickness days in 1999

back pain.[6] These constitute more than 80% of the total sample, and we have generated dummy variables for these two groups. The third medical group consists of patients with back pain-related diagnoses other than disk herniation and low back pain. This group is used as the base category in the regressions. In addition, we have information about annual income in 1999. We also use the number of sickness days prior to the period of sickness in 2000. However, we have constucted the data set in such a way that no patients were on sick leave during the 6-month period prior to taking sick leave in 2000. Thus, all patients in our data set could have been on sick leave for up to 12 months after starting their sick leave in 2000.

From information on where people live, we have constructed a variable that measures the distance in kilometres to the nearest hospital that offers a treatment programme for these kinds of patients. As is explained in Section 5, this variable is the excluded variable in the model that controls for selection into treatment. Our hypothesis is that there is a higher probability of getting treatment if the patient lives close to a hospital that offers treatment. This variable should not affect the transition out of sickness benefits, except indirectly through treatment.

Descriptive statistics for the full sample are reported in Table 2a, while Table 2b and c describe the samples for the treated and non-treated patients, respectively. Interestingly, we find that the proportion leaving the sickness benefits scheme after 9 months is greater for treated than non-treatment patients (0.481 versus 0.408). The unconditional means differ by 7.3 and this difference would have been a consistent measure of the treatment effect if our data were truly experimental. However, our data are observational, and the two samples are therefore potentially unbalanced both in observables and unobservables. This could be illustrated by looking at the diagnosis variables, for example. We see that the proportion of patients with a diagnosis of disk herniation is much higher in the non-treatment group than in the treatment group, while the proportion with low back pain is highest in the treatment group. As expected, we find that the distance to the nearest hospital that offers treatment

[6] Low back pain, or non-specific low back pain, is a symptom that can have many causes. Many cases of back pain are caused by stress on the muscles and ligaments that support the spine. Both increased weight on the spine and increased pressure on the discs can cause low back pain. A low back problem may come on suddenly or gradually.

Table 2
Descriptive statistics

Variable	Mean	S.D.	Minimum	Maximum
(a) Full sample, $n = 656$				
Distance	4.092	1.554	0.0	7.6
Male	0.587	0.493	0.0	1.0
Age	41.380	10.962	19.0	67.0
Age^2	1832.245	916.367	361.0	4489.0
Disk herniation	0.305	0.461	0.0	1.0
Low back pain	0.495	0.500	0.0	1.0
Income 1999	221.942	111.283	9.7	1328.1
Sick days 1999	41.800	62.184	0.0	309.0
Treatment	0.159	0.366	0.0	1.0
Out of sickness	0.419	0.494	0.0	1.0
(b) Treated, $n = 104$				
Distance	2.524	1.564	0.0	5.6
Male	0.567	0.498	0.0	1.0
Age	41.067	10.310	20.0	63.0
Age^2	1791.798	833.657	400.0	3969.0
Disk herniation	0.135	0.343	0.0	1.0
Low back pain	0.567	0.498	0.0	1.0
Income 1999	203.999	80.146	9.7	500.1
Sick days 1999	42.904	69.136	0.0	283.0
Treatment	1.000	0.000	1.0	1.0
Out of sickness	0.481	0.502	0.0	1.0
(c) Non-treated, $n = 552$				
Distance	4.388	1.365	0.0	7.6
Male	0.590	0.492	0.0	1.0
Age	41.438	11.088	19.0	67.0
Age^2	1839.866	931.634	361.0	4489.0
Disk herniation	0.337	0.473	0.0	1.0
Low back pain	0.482	0.500	0.0	1.0
Income 1999	225.322	115.967	10.7	1328.1
Sick days 1999	41.592	60.852	0.0	309.0
Treatment	0.000	0.000	0.0	0.0
Out of sickness	0.408	0.492	0.0	1.0

(distance) is highest among individuals in the non-treatment group. For the variables male, age, number of sickness days in 1999, and income in 1999, there are only small differences between the two samples. To take the non-experimental nature of our data into consideration, we analyse the selection process into treatment formally using the econometric model outlined in Section 2.

4. Selection into treatment

We first discuss the parameters relating to selection into treatment within the framework of a regression model. The selection parameters reported in Table 3 offer a straightforward

Table 3
Probit model of treatment decision

Treatment	Coefficient	S.E.	$P > \|z\|$	95% CI	dF/dx
Distance	−0.43137	0.04453	0.000	[−0.5186, −0.3440]	−7.64215
Male	0.07167	0.14544	0.622	[−0.2133, 0.3567]	1.25985
Age	0.06944	0.04789	0.147	[−0.0244, 0.1633]	1.23025
Age^2	−0.00080	0.00057	0.160	[−0.0019, 0.0003]	−0.01433
Disk herniation	−0.59433	0.20457	0.004	[−0.9952, −0.1933]	−9.14283
Low back pain	−0.38125	0.16302	0.019	[−0.7007, −0.0617]	−6.76448
Income 1999	−0.00192	0.00085	0.025	[−0.0036, −0.0002]	−0.03409
Sick days 1999	0.00123	0.00103	0.234	[−0.0007, 0.0032]	0.02180
Constant	−0.19691	0.96214	0.838	[−2.0826, 1.6888]	

Number of observations = 656, LR $\chi^2(8) = 137.42$, pseudo-$R^2 = 0.2396$, Prob $> \chi^2 = 0.0000$, log likelihood = −218.11563.

way of examining the presence of non-random selection into treatment. Table 3 presents the estimated coefficients of the probit model. Several of the estimated coefficients are statistically different from zero, as can be seen from the z-values. This indicates that individuals under treatment differ significantly from eligible non-participants with respect to observable characteristics. The likelihood ratio test also concludes that the treatment group and the comparison group differ. The last column in Table 3 shows the marginal effects in percent.

Table 3 shows that the effect of distance to the nearest hospital (our instrument) is significantly different from zero. Thus, the first requirement of having a valid instrument, that it is correlated with the treatment decision, is satisfied. The instrument should however not affect the outcome directly, only indirectly through the treatment variable.

The effects of age and gender are not significant in the selection equation. However, the effects of both medical dummies are different from zero. The effect of income is also significantly different from zero. Higher income in 1999 reduces the probability of being treated at an outpatient hospital. Table 4 shows the number of correct predictions in the probit model. The model fit is relatively good. The pseudo-R^2 reported in Table 3 is 0.24. The support of the predicted probabilities for treatment is very good given the relatively low number of observations in the data set. The support region for the propensity score is [0.024, 0.820] for participants and [0.0001, 0.850] for non-participants, with a mean of 0.34

Table 4
Number of correct predictions in selection equation

Treatment	Treatment predicted		Total
	0	1	
0	427	125	552
1	22	82	104
Total	449	207	656

Table 5
Discrete-choice model of leaving the sickness benefit scheme

Out of sickness	Participation state		Non-participation state	
	No selection	Selection	No selection	Selection
Male	0.01938 (0.27945)	0.02289 (0.28153)	0.17069 (0.11966)	0.16186 (0.11956)
Age	−0.19572* (0.10140)	−0.19577* (0.08823)	0.03034 (0.03773)	0.02219 (0.03826)
Age^2	0.00207* (0.00123)	0.00208* (0.00107)	−0.00060 (0.00045)	−0.00050 (0.00046)
Disk herniation	0.14124 (0.42629)	0.17920 (0.43330)	0.13660 (0.16156)	0.20465 (0.16752)
Low back pain	−0.32876 (0.29902)	−0.31343 (0.29783)	−0.07086 (0.15331)	−0.04748 (0.15184)
Income 1999	0.00344* (0.00187)	0.00337* (0.00192)	0.00144** (0.00056)	0.00154** (0.00052)
Sick days 1999	0.00180 (0.00191)	0.00188 (0.00191)	0.00252** (0.00091)	0.00246** (0.00089)
Constant	3.66890* (1.96800)	3.77682* (1.75807)	−0.94462 (0.75837)	−0.89665 (0.76299)
ρ		−0.11045 (0.28650)		0.30691 (0.25486)
α		−0.15814		0.48178

We have used the Huber/White/sandwich estimator of the variance. S.E. given in parentheses.

and 0.12, respectively.[7] The support region is larger for non-participants, which could be explained by the relatively large number of non-participants (552) compared to the number of participants (104).

5. Transition out of sickness benefits

We are interested in the effect of treatment on leaving the sickness benefits scheme after 9 months. We first look at the outcome model that does not include unobserved heterogeneity. We report the estimated outcome regression coefficients in Table 5, where the β_1-vector without selection is reported in column 1 of Table 5, and the β_0-vector without selection is reported in column 3 of Table 5. For both outcome equations, it is surprising that variables such as gender, age, and diagnosis do not have statistically significant effects on the probability of leaving the sickness benefits scheme. However, the income variable and the number of sickness days in 1999 are statistically significant. Higher income increases the probability of leaving the sickness benefits scheme. A higher number of sickness days in 1999 also increases the probability of returning to work.

For the model with no unobserved heterogeneity, if we condition on a given X-value, the average treatment effect and the effect of treatment on the treated parameters are equivalent. However, if we average over different distributions of X to get the unconditional average treatment effect and the effect of treatment on the treated (that is, average over the unconditional distribution of X for the average treatment effect and average over the distribution of X conditional on $D = 1$ for the effect of treatment on the treated), the resulting averaged version of these parameters will be different. Using the results of our model without unobserved heterogeneity, the estimated average effect of treatment (averaging over the

[7] Any non-experimental evaluation can non-parametrically estimate treatment effects only over the common support region: see Heckman et al. (1998). Due to the relatively low number of observations, we do not pursue non-parametric estimation of treatment effects. For a non-parametric matching strategy: see Aakvik (2001).

unconditional empirical distribution of X) is 12.3 percentage points. The estimated average effect of treatment on those treated (averaging over the empirical distribution of X conditional on $D = 1$) is 9.3 percentage points.

Persons treated have observable characteristics that are associated with a slightly lower effect of treatment, so that on average, their treatment effect is lower than it would be for a person drawn at random from the pool of patients who have been examined at an outpatient clinic. The unconditional mean difference of leaving the sickness benefits scheme between treated and non-treated, when we do not control for observed background variables, is 7.3 percentage points. Thus, the treatment effect increases once we adjust for observed variables, and it increases more for a random patient in the sample.

The model with unobserved heterogeneity allows selection both on observables and unobservables. We have used the distance to nearest hospital that treats patients as our identifying exclusion restriction in the selection model. Distance from hospital represents an extra cost of treatment for the patient. The further away from the hospital the patient lives, the higher this cost will be. We can see from Table 3 that the effect of this variable is significant in the decision to be treated. However, for this variable to be a valid instrument, it should not affect the probability of leaving the sickness benefits scheme, except indirectly through the treatment variable. The estimated coefficient of the distance variable was close to zero with a z-value of 0.1 when included in the outcome equations, while it was highly statistically significant in the selection equation with a z-value of almost 10. Thus, there is no indication that distance to hospital affects the transition out of the sickness benefits scheme. Geographic variation is an often used instrument in the literature on the returns to education: see, for instance, Card (1995), where it affects the probability of entering college but not subsequent earnings.

Column 2 of Table 5 shows the estimated parameter vector in the sickness outcome for treated patients (β_1), while column 4 of Table 5 shows the results for non-treated patients (β_0), for the model with unobserved heterogeneity correlated over states. The effect of treatment on the treated patients drops to 6.3 percentage points compared to the unconditional effect of 7.3 percentage points. Although the factor structure model estimates insignificant factor loadings, the selection specification still affects the estimated effect of treatment. In fact, it reduces the treatment effect. This suggests that the unobserved elements of selection into treatment reduces the treatment effect. This is caused by the fact that α_1 is negative and α_0 is positive, although neither are significantly different from zero.

Fig. 1 shows the distribution of TT within the random coefficient framework. The distribution of TT is slightly right-skewed. The maximum estimated treatment effect is 50 percentage points. However, the majority of the treated patients, have a treatment effect in the interval $[-0.1, 0.3]$. The mean treatment effect is 6.3 percentage points.[8] Fig. 1 also indicates that treatment of low back pain patients can be counterproductive, that is, the model predicts a negative treatment effect for some patients, but the overall effect is positive.

Given this positive effect of treatment, it might be of some interest to illustrate the impact of expanding the programme of multidisciplinary outpatient treatment. The treatment effect in percentage-point terms can be translated into a number of days. We find a treatment effect

[8] As pointed out in Section 3, the Norwegian sickness benefits system is quite generous and workers lack financial incentives to leave the sickness benefits scheme after successful treatment. Thus, the estimated treatment effect might be under-estimated compared to a less generous system.

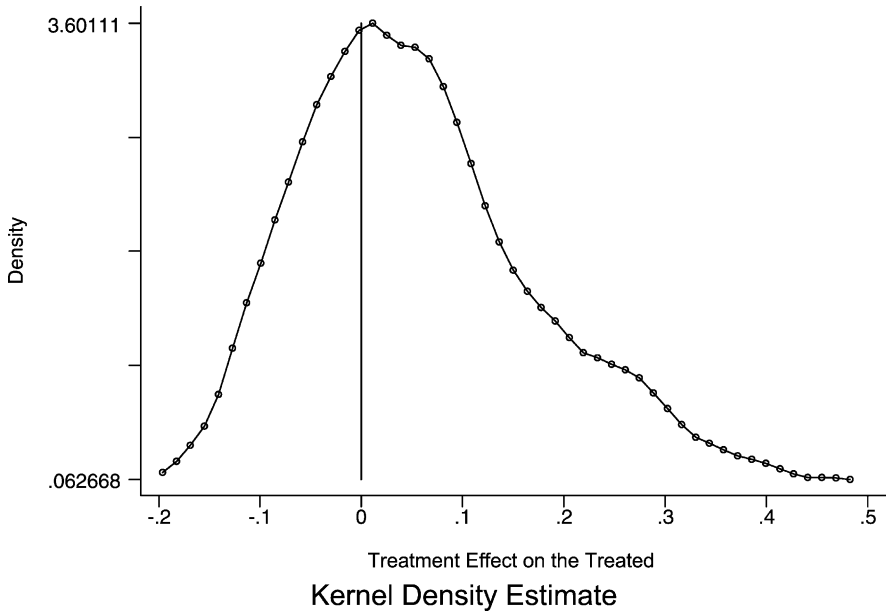

Fig. 1. The distribution of TT with unobserved selection (mean TT = 6.3 percentage points) (vertical line is 0 treatment effect).

for an average patient of around 8.5 days, adjusted for holidays, based on the treatment effect of 6 percentage points. Our data only cover patients examined at outpatient spinal clinics. Hagen and Thune (1998) identify a total of 89,190 patients with low back pain who received sickness benefits for at least 2 weeks in 1995 and 1996 (about 45,000 patients each year). However, not all patients would benefit equally from treatment. Haldorsen et al. (2002) categorize back pain patients into three groups differing in prognosis score (good, medium, and poor) for return to work by a screening instrument. Of 654 patients, 22% are found to have good prognosis, 50% to have medium prognosis, and 28% to have poor prognosis. Compared to ordinary treatment, Haldorsen et al. (2002) only found treatment effects from multidisciplinary treatment for patients with a poor or medium prognosis. This indicates that maximum 78% of the back pain patients in Norway could potentially benefit from receiving multidisciplinary treatment each year. Using the results of Hagen and Thune (1998), we estimate the number of patients suitable for treatment to be around 35,000.[9] Assuming that

[9] Note that Haldorsen et al. (2002) only include patients on long-term sick leave (sick-listed for more than 8 weeks), while Hagen and Thune (1998) include patients who are sick-listed for at least 2 weeks. The proportion of patients with good prognoses is probably higher in the sample used by Hagen and Thune (1998). In addition, the median duration of sickness spells is 43 days in Hagen and Thune (1998). In our sample, the duration of absence is much higher. The average sickness period in our data is 222 days. This probably means that the number of patients having a good prognosis is relatively higher in the entire population of back pain patients than is the case in our sample. By extrapolating our results to the entire population, therefore, we overestimate the number of patients that could potentially be treated. However, even if we assume a lower number, the effect on the sickness benefits scheme is substantial.

all patients suffering from back pain have the same effect of treatment as the patients in our sample, our best estimate is that a programme expansion could save a maximum of 300,000 sick-leave days each year.

6. Conclusions

In this paper, we have estimated treatment effects for back pain patients using observational data. We find a positive effect of treatment of 6 percentage points on the probability of leaving the sickness benefits scheme and returning to work after allowing for selection effects and full heterogeneity in treatment effects. We also find that there are sound arguments for increasing the outpatient programme of treating back pain patients. Usually, the average treatment effect is expected to be lower than the effect of treatment on the treated if individuals are rational and can act on the unobserved element of selection on the transition out of sickness benefits. We do not find any significant unobserved selection effects in this data set. This may be explained by the fact that it is difficult for individuals to predict the outcome of treatment of back pain diagnoses. The suggestion that there are sound arguments for programme expansion should be subject to the qualification that we have not considered the cost of treatment or investment costs in this paper. However, both Skouen et al. (2002) and Haldorsen et al. (2002) find that treatment for certain patients is cost effective. Reviewing the international literature, Guzmán et al. (2001) also find positive effects on such medical outcomes as pain and function, but find less clear results on vocational outcomes.

In addition to treatment costs and investment costs, an expansion of the multidisciplinary treatment programme for low back pain patients may incur additional costs. In our study, we experienced that many hospitals that wanted to implement multidisciplinary treatment lacked medical personnel. In Norway, it is not rare that lack of medical personnel has turned out to be a bottleneck when hospitals have tried to expand treatment programmes. Thus, giving strong incentives to increase capacity may have a negative spillover effect on other treatment programmes at the hospitals. Although our estimates of treatment effects are positive, and although we find that a substantial number of sickness days could be saved in the sickness benefits scheme, a more thorough discussion of programme expansion must include cost estimates.

Acknowledgements

We thank the Norwegian Ministry of Health for financial support. We gratefully acknowledge comments from Christoph Schmidt, Søren Brage, participants at workshops in Lund and Bonn, and two anonymous referees.

References

Aakvik, A., 2001. Bounding a matching estimator: the case of a Norwegian training program. Oxford Bulletin of Economics and Statistics 63 (1), 115–143.

Aakvik, A., Heckman, J.J., Vytlacil, E.J., 2000. Treatment effects for discrete outcomes when responses to treatment vary among observationally identical persons: an application to Norwegian vocational rehabilitation programs. NBER Technical Working Paper, 262.

Butler, J., Moffitt, R., 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. Econometrica 50 (3), 761–764.

Card, D., 1995. Using geographic variation in college proximity to estimate the return to schooling. In: Christofides, L., Grant, E., Swidinsky, R. (Eds.), Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp. University of Toronto Press, Toronto, pp. 201–222.

Cox, D.R., 1958. The Planning of Experiments. Wiley, New York.

Fisher, R.A., 1935. Design of Experiments. Oliver and Boyd, London.

Guzmán, J., Esmail, R., Karjalainen, K., Malmivaara, A., Irvin, E., Bombardier, C., 2001. Multidisciplinary rehabilitation for chronic low back pain: systematic review. British Medical Journal 322, 1511–1516.

Hagen, K.B., Thune, O., 1998. Work incapacity from low back pain in the general population. Spine 23 (19), 2091–2095.

Ham, J.C., LaLonde, R.J., 1996. The effects of sample selection and initial conditions in duration models: evidence from experimental data. Econometrica 64 (1), 175–205.

Haldorsen, E.M.H., Grasdal, A.L., Skouen, J.S., Risa, A.E., Kronholm, K., Ursin, H., 2002. Is there a right treatment for a particular group? Pain 95, 49–63.

Heckman, J.J., 1978. Endogenous variables in a simultaneous equation system. Econometrica 6 (4), 931–959.

Heckman, J.J., 1981. Statistical models for discrete panel data. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, New York, pp. 114–178.

Heckman, J.J., 1997. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. Journal of Human Resources 32, 441–462.

Heckman, J.J., Honoré, B., 1990. The empirical content of the Roy model. Econometrica 58, 1121–1149.

Heckman, J.J., Robb, R., 1985. Alternative methods for evaluating the impact of interventions. In: Heckman, J.J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data. Cambridge University Press, New York, pp. 156–245

Heckman, J.J., Vytlacil, E.J., 2000. The relationship between treatment parameters within a latent variable framework. Economics Letters 66, 33–39.

Heckman, J.J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing selection bias using experimental data. Econometrica 66, 1017–1098.

Indahl, A., Velund, L., Reikerås, J., 1995. Good prognosis for low back pain when left untreated. Spine 4, 473–477.

Karjalainen, K., Malmivaara, A., van Tulder, M., Roine, R., Jauhiainen, M., Hurri, H., Koes, B., 2001. Multidisciplinary biopsychosocial rehabilitation for sub-acute low back pain in working-age adults: a systematic review within the framework of the Cochrane Collaboration Back Review Group. Spine 26 (3), 262–269.

Neyman, J., 1923. Statistical problems in agricultural experiments. Journal of the Royal Statistical Society 2 (2), 107–180.

Nordic Social-Statitical Commitee (NOSOSCO), 2001. Social Protection in the Nordic Countries 1999. NOSOSCO 16:2001, Copenhagen.

Quandt, R., 1972. Methods for estimating switching regressions. Journal of The American Statistical Association 67 (338), 306–310.

Roy, A., 1951. Some thoughts on the distribution of earnings. Oxford Economic Papers 3, 135–146.

Rubin, D., 1978. Bayesian inference for causal effects: the role of randomization. Annals of Statistics 6, 34–58.

Skouen, J.S., Grasdal, A.L., Haldorsen, E.M.H., Ursin, H., 2002. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine 27 (9), 901–910.

van Tulder, M.W., Ostelo, , R, , Vlaeyen, J.W.S., Linton, S.J., Morley, S.J., Assendelft, W.J.J., 2001. Behavioral treatment for chronic low back pain—a systematic review within the framework of the Cochrane Back Review Group. Spine 26 (3), 270–281.