A maximum likelihood estimator of a Markov model for disease activity in Crohn's disease and ulcerative colitis for annually aggregated partial observations.

Borg, Sixten; Persson, Ulf; Jess, Tine; Østergaard Thomsen, Ole; Ljung, Tryggve; Riis, Lene; Munkholm, Pia

Published in:
Medical Decision Making

DOI:
10.1177/0272989X09336141

2010

A Maximum Likelihood estimator of a Markov model for disease activity in Crohn's disease and ulcerative colitis for annually aggregated partial observations.

Sixten Borg, M.Sc. (1), Ulf Persson, Ph.D. (1), Tine Jess, MD (2), Ole Østergaard Thomsen, MD, DMSci (3), Tryggve Ljung, MD, PhD (4), Lene Riis, MD (3), Pia Munkholm, MD, DMSci (3).

(1) IHE, The Swedish Institute for Health Economics, Lund, Sweden

(2) Institute of Preventive Medicine, Copenhagen University Hospital, Copenhagen, Denmark

(3) Danish Crohn's Colitis Database, Department of Medical Gastroenterology, Herlev University Hospital, Copenhagen, Denmark.

(4) Karolinska University Hospital, Stockholm, Sweden,

2008-12-03

Corresponding author: Sixten Borg, The Swedish Institute for Health Economics (IHE), Box 2127, S-220 02 Lund, Sweden. Phone: +46 46 32 91 00. Fax: +46 46 12 16 04.

Abstract

Crohn's disease (CD) and ulcerative colitis (UC) are chronic inflammatory bowel diseases that have a remitting, relapsing nature. During relapse, they are treated with drugs and surgery. The present study was based on individual data from patients diagnosed with CD or UC at Herlev University Hospital, Copenhagen, Denmark, during 1991-1993. The data was aggregated over calendar years, for each year the number of relapses and the number of surgical operations were recorded. Our aim was to estimate Markov models for disease activity in Crohn's disease and ulcerative colitis, in terms of relapse and remission, with a cycle length of one month. The purpose of these models was to enable evaluation of interventions that would shorten relapses or post-pone future relapses. An exact maximum likelihood estimator was developed, that disaggregates the yearly observations into monthly transition probabilities, between remission and relapse. These probabilities were allowed to be dependent on the time since start of relapse, and on the time since start of remission, respectively. The estimator, initially slow, was successfully optimized to shorten the execution time. The estimated disease activity model for Crohn's disease fits well to observed data and has good face validity. The disease activity model is less suitable for ulcerative colitis, due to its transient nature through the presence of curative surgery.

Key words: maximum likelihood estimator, aggregated and partial observations, Markov model, transition probability matrix, inflammatory bowel disease, disease activity.

1. Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are chronic inflammatory bowel diseases that have a remitting, relapsing nature. Relapses are manifested as increased inflammatory activity and increased symptoms such as abdominal pain, fever, and weight loss [1, 2]. During a relapse, CD and UC are treated with drugs such as oral and topical glucocorticosteroids, 5-aminosalicylates (including sulphasalazine), immunosuppressive drugs (e.g. azathioprine, 6-mercatopurines, methotrexate, cyclosporine), antibiologics (infliximab), and antibiotics, or with surgery [1, 2]. In CD, the most common surgical procedure is resection of a part of the intestine that is severely affected by the disease, in the lack of response to drug therapy [1,3,4]. Resection is not a curative treatment and may need to be performed many times, since the disease may reappear elsewhere in the intestinal tract [1]. In UC, which by definition only involves colon, surgery may be considered curative if the entire colon is removed by colectomy and restorative surgery such as ileoanal pelvic pouch (IAPP) or ileorectal anastomose (IRA) is performed [1,4,5]. After such surgery, UC cannot relapse again. However, the patient may suffer complications following IAPP or IRA, namely inflammation in the pouch or rectum (pouchitis, IRA-proctitis) [4].

An intervention aimed at shortening relapses will aim to increase the probability of going from relapse to remission, and interventions to post-pone the next relapse (i.e. a more successful maintenance treatment) will decrease the probability of going from remission into relapse. In order to analyse such interventions, a model of remission and relapse is needed. Our aim was to estimate the parameters of Markov models for CD and UC patients that alternate between remission and relapse. In addition, the model needed to run in reasonably short cycles to allow us to study the effect of shortening the relapses.

Our data were partial observations, aggregated over one-year periods. A translation from the one-year perspective into a one-month perspective was required. Translating from long to short intervals of time is fairly simple for a single risk [6]. It is more complicated for transition probabilities in models with several states that allows transition back and forth, i. e. transition probability matrices [6,7,8,9]. One such method is matrix decomposition which defines the short-term matrix as a function of the long-term matrix. It determines the short-interval matrix in closed form, i. e. using a single formula [8,10]. However, in some circumstances the short-interval matrix is an invalid transition probability matrix (e. g. it may contain negative values).

Craig & Sendi [8] demonstrate how the Expectation-Maximization (E-M) algorithm can be used to approximate the Maximum Likelihood (ML) in a situation with partial observations, where the algorithm is used to impute data at time points not observed. They demonstrate how this overcomes the problem of invalid short-term matrices. Charitos et al [10] employ matrix decomposition as well, and they use regularization techniques to deal with invalid short-term matrices. They use their technique to refine an invalid matrix until they find the nearest valid matrix, and use that one instead.

Welton and Ades [9] use a Bayesian approach in a similar situation. They present how to estimate transition rates for multi-state models from partial data, which consist of observations of the initial state and the state after a certain period of time. Their data are aggregated in the sense that they look at transition counts from groups of patients rather than individual patient data. Once having estimated the transition rates, the transition probabilities are easily obtained for any desired time interval.

Craig, Fryback et al [11] have observations at uneven intervals (year 1, 5 and 11), of patients with diabetes, where the patients are classified into different stages of retinopathy. Some of the patients are given a treatment intervention at some stage within this range of years. The authors give the transition probabilities a parametric form and in addition, they add parametric forms for the treatment intervention and for death. Hereby they manage to estimate the natural history of diabetic retinopathy using a bayesian approach employing a Markov Chain Monte Carlo (MCMC) technique [11].

These approaches have in common that they have observations of the process at long intervals, and tries to determine what occurs at time points in between. They all have data on visits to, or transitions between, the states that are of interest.

This is different from our situation, as we neither have observations of transitions nor of the patient's state at any point in time. Instead, we have yearly counts of events, namely the number of relapses, and the number of surgical operations, for each patient. We do not know whether they are in relapse or remission at a given point in time, except that they are initially in relapse. Thus our data are not aggregated data from several patients. In our case, each individual patient's events during a year's time is aggregated into summary counts, describing the whole year, for that individual patient. We use the Markov model itself as a framework to produce an estimator of the sought parameters, using the Maximum Likelihood (ML) method. The ML method estimates its parameter at the value that maximizes the likelihood of observing the actual observations [12].

2. Material

The present study was based on data from patients diagnosed with CD or UC at Herlev University Hospital, Copenhagen, Denmark, during 1991-1993 (described in detail elsewhere) [2]. The data was organized in calendar years. For each year, the number of relapses and the number of surgical operations were recorded. A relapse is defined as a contact with a physician due to gastrointestinal symptoms, that led to increased use of medication or surgery [2]. The year was classified according to a disease course level: *silent, mild* and *severe*. A silent year is a year completely in remission. A mild year has 1-2 relapses but no surgery. A severe year has more than two relapses, or presence of surgery, or both. This is a simplified definition of disease course which aims to resemble earlier definitions of annual classifications of disease activity [3,5], and which will be clinically correct for most situations. There are clinical situations where a discrepancy may occur, e. g. a whole year with continuous severe symptoms would be classified as a mild year with one relapse here, but would be regarded as a severe year clinically. The types of surgery recorded in the database were total and subtotal colectomy, small intestine resection, colon resection and "other", such as fistula surgery or stricture plastic.

A total of 145 individuals, of which 58 were diagnosed with CD and 87 with UC, were observed for a total of 1,292 patient-years (Table 1). The CD patients had average rates of 0.60 relapses per year (range, 0 - 5 relapses) and 0.24 surgical operations per year (range, 0 – 5 procedures) (Table 2). The UC patients had 0.69 relapses/year (range, 0 - 7) and 0.04 surgical operations/year (range, 0 - 4).

3. Methods

Our data consists of yearly observations of the number of relapses and the number of surgical operations that the patient has experienced within the year. The notation used to describe the data is as follows. We observe an individual i for n_i years; its number of surgical operations $Z_i = (Z_{i1} \quad Z_{i2} \quad ... \quad Z_{in_i})$, and its number of relapses $V_i = (V_{i1} \quad V_{i2} \quad ... \quad V_{in_i})$, during each observed year. As an example, one particular patient was observed during 11 years, of which the two first contained one relapse each. During the second year, the patient was subjected to two surgical operations. The disease was silent from the third year and on, with no new relapses or surgery. Thus, $n_i = 11$, $V_i = (1,1,0,0,0,0,0,0,0,0,0)$, and $Z_i = (0,2,0,0,0,0,0,0,0,0,0)$.

A Markov model was constructed for the disease activity in CD and UC. We developed an exact Maximum Likelihood (ML) estimator for the parameters of this model [12,13]. The ML estimator uses the probability of observing the actual observations, conditionally on the parameter vector. By selecting the parameter values so that the probability is maximized, estimates of the parameters are obtained. Sometimes, it is possible to derive the estimator in closed form. Otherwise, the parameter values are found using a numerical search. To avoid finding just local maxima, several starting points are used. We used 20 starting points. The estimator was used to transform the yearly data into monthly probabilities. We tried to determine whether the estimated disease activity models could successfully predict disease course, and whether the estimator appeared to work and how accurate its estimates were. The model, estimator and our approach to judge it's success are described in the following sections.

3.1 The disease activity model

The disease activity model is a Markov chain, $S_t, t = 0,1,...$ with four states, first month of remission ($S_t = 1$), subsequent months of remission ($S_t = 2$), first month of relapse ($S_t = 3$), and subsequent months of relapse ($S_t = 4$), which is presented in Figure 1. The transition probability matrix is

$$P = \begin{pmatrix} 0 & 1-p_3 & p_3 & 0 \\ 0 & 1-p_4 & p_4 & 0 \\ p_1 & 0 & 0 & 1-p_1 \\ p_2 & 0 & 0 & 1-p_2 \end{pmatrix}$$, where p_1, p_2, p_3,and p_4, are the transition probabilities (p_1 from state 3 to 1, p_2 from state 4 to 1, p_3 from state 1 to 3, and p_4 from state 2 to 3). These four transition probabilities and p_5, the probability of surgery while visiting state 3 or 4, are the five parameters to the model which we aim to estimate. Time dependence is restarted for each new visit to remission and relapse. The relapse states form a two-state tunnel. A consequence of this construction is that we can easily model a mixture of short and long visits to e. g. relapse, say $p_1 = 1/2$, $p_2 = 1/5$ would make half the visits one cycle long, and the remainder (conditional on having stayed the first month) geometrically distributed with a mean duration of 5 cycles, i.e. 5 months. The remission states form the same construction, with it's own parameters. Our model is the minimal Markov model that has time-dependent probabilities of changing from remission to relapse, and vice versa. The probability of relapse (and remission, respectively) is dependent on the time since entry into remission (and into relapse, respectively) as this is embedded in the model structure, but the Markov model itself is time-homogeneous, since no parameter value (p_1, p_2, ..., p_5) will change over time.

Along with the Markov chain, $S_t, t = 0,1,...$, we use a surgery indicator $T_t, t = 0,1,...$ whether the patient was subjected to surgery while visiting state 3 or 4, i. e. $\Pr\{T_t = 1\} = p_5$ if $S_t = 3$ or $S_t = 4$, and 0 otherwise.

The disease course of a twelve-month sequence of disease activity $\{(S_t, T_t)\}, t = 1,2,...,12$ is determined according to the definition given above. If $\{T_t = 1\}$ for any $t = 1,2,...,12$, or if there are three or more visits into the relapse states, the disease course is *severe*. If there are one or two visits into the relapse states and $\{T_t = 0\}$ for all $t = 1,2,...,12$, the disease course is *mild*. In case of no visits to the relapse states, the disease course is *silent*.

A relapse in the model begins with a transition into state 3, possibly followed directly by one or more cycles spent in state 4. The relapse ends with a transition to state 1. A relapse beginning one year and lasting into the next year contributes to the relapse counts in both years.

The disease activity Markov chain has the following stationary distribution;

$$\pi_1 = \pi_3 = \frac{1}{2p_2p_4 + p_2(1-p_3) + p_4(1-p_1)},\ \pi_2 = \frac{1-p_3}{p_4}\pi_1, \text{ and } \pi_4 = \frac{1-p_1}{p_2}\pi_1,$$

where $p_2 > 0, p_4 > 0$. We assume that this stationary distribution and our disease activity model are relevant during the period when the patients were observed.

3.2 The exact Maximum Likelihood estimator

The likelihood for an individual i is $L_i(\theta) = \Pr\{Z_i = z_i, V_i = v_i \mid \theta\}$, a function of the disease activity model parameters $\theta = (p_1 \quad p_2 \quad p_3 \quad p_4 \quad p_5)$. The probability of the observations

during a given year l and the disease activity as the next year begins (S_{13}), conditionally on the disease activity at its beginning (S_1) and θ, is denoted

$$y_{s_1,s_{13}}(z_l, v_l) = \Pr\{Z_l = z_l, V_l = v_l, S_{13} = s_{13} \mid S_1 = s_1, \theta\} =$$

$$= \sum_{s_2,s_3,\cdots,s_{12}=1}^{4} \left(\sum_{t_1,t_2,\cdots,t_{12}=0}^{1} I_1(\tilde{t}) \cdot I_2(\tilde{s}) \cdot \Pr\{S_2 = s_2, S_3 = s_3, \cdots, S_{13} = s_{13}, T_1 = t_1, T_2 = t_2, \cdots, T_{12} = t_{12} \mid S_1 = s_1, \theta\} \right)$$

, where $I_1(\tilde{t}) = I\{t_1 + t_2 + \cdots + t_{12} = z_l\}$, and $I_2(\tilde{s}) = I\{v(s_1, s_2, \cdots, s_{12}) = v_l\}$.

The summation goes over all possible sequences of $\{S_r\}_{r=2}^{12}$ and $\{T_r\}_{r=1}^{12}$. The indicators select those sequences, which have z_l surgical operations and v_l relapses, respectively. The indicator of sequences that have v_l relapses cannot be given explicitly, but it indicates whether the number of visits into the group of states 3 and 4 as described above is equal to v_l. The last term is the probability of the sequences $\{S_r\}_{r=2}^{12}$ and $\{T_r\}_{r=1}^{12}$, and this probability is a product of elements in the disease activity transition probability matrix (P_{s_1,s_2} is the one-month probability of transition from s_1 to s_2), and of the probabilities of a set of Bernoulli trials,

(eq. 1) $$\Pr\{S_2 = s_2, S_3 = s_3, \cdots, S_{13} = s_{13}, T_1 = t_1, T_2 = t_2, \cdots, T_{12} = t_{12} \mid S_1 = s_1, \theta\} =$$

$$= P_{s_1,s_2} \cdot P_{s_2,s_3} \cdot \cdots \cdot P_{s_{12},s_{13}} \cdot p_5^{\sum I\{t_i=1, s_i \in \{3,4\}\}} (1 - p_5)^{\sum I\{t_i=0, s_i \in \{3,4\}\}}.$$

This probability will become zero for many of the sequences S_t whenever one or more of the corresponding transition probabilities (P_{s_1,s_2}) are zero.

Then we can write the likelihood for an individual's observed sequence by counting over all possible disease activity ($S_{i,l,1}$) at the beginning of each year, $l = 1,2,\cdots,n_i + 1$ (including the final target state);

$$L_i(\theta) = \sum_{s_{i,1,1},s_{i,2,1},\cdots,s_{i,n_i+1,1}=1}^{4} y_{s_{i,1,1},s_{i,2,1}}(z_{i1},v_{i1}) y_{s_{i,2,1},s_{i,3,1}}(z_{i2},v_{i2}) \cdots y_{s_{i,n_i,1},s_{i,n_i+1,1}}(z_{in_i},v_{in_i}) \Pr\{S_{i,1,1} = s_{i,1,1}\}.$$

The first part is a series of matrix multiplications. The last factor, the probability of the initial state, is taken from the stationary distribution π. Since we know that the patients start with a relapse, $S_{i,1,1}$ must be 3 or 4. Conditionally on this, we get that $\Pr\{S_{i,1,1} = j\} = \pi_j / (\pi_3 + \pi_4)$.

A maximum likelihood estimate was sought by maximizing the product of all the individuals' likelihoods $L(\theta) = \prod_i L_i(\theta)$ with regards to θ. This process is very time-consuming, since the probabilities $y_{s,u}(z,v)$ were computed exactly by traversing every possible sequence of disease activity.

However, the structure of the model never changes. Each possible pathway through the model is a product of the probabilities p_1, p_2, ..., p_5 and their complementary probabilities. Therefore, equation 1 can be rewritten using a set of exponents $a_1, b_1, \cdots, a_5, b_5$,

$$P_{s_1,s_2} \cdot P_{s_2,s_3} \cdot \cdots \cdot P_{s_{12},s_{13}} \cdot p_5^{\sum I\{t_i=1, s_i \in \{3,4\}\}} (1-p_5)^{\sum I\{t_i=0, s_i \in \{3,4\}\}} = (p_1)^{a_1}(1-p_1)^{b_1}(p_2)^{a_2}(1-p_2)^{b_2} \cdots (p_5)^{a_5}(1-p_5)^{b_5}.$$

The exponents a_1, b_1, a_2, b_2, a_3, b_3, a_4, and b_4, are the number of times p_1, $(1-p_1)$, p_2, $(1-p_2)$, p_3, $(1-p_3)$, p_4, and $(1-p_4)$, respectively, occur as factors in the probability

expression. They are determined for each pathway through the model. The last two exponents are taken directly from eq. 1, $a_5 = \sum I\left\{t_i = 1, s_i \in \{3,4\}\right\}$, and $b_5 = \sum I\left\{t_i = 0, s_i \in \{3,4\}\right\}$.

Not all sequences S_t and T_t generate unique sets of exponents. We define a *profile* to be the set of numbers $s_1, s_{13}, a_1, b_1, a_2, b_2, \cdots, a_5, b_5, v, r$, which describe the r sequences with v relapses and a_5 surgical operations, that begin with $S_1 = s_1$, and end with $S_{13} = s_{13}$, and that all have a probability that can be described with the exponents $a_1, b_1, \cdots, a_5, b_5$ as shown above. Reducing the pathway to such exponents results in loss of information on the number of relapses. This is therefore recorded in the profile as v.

Therefore we can determine $y_{s_1,s_{13}}(z,v)$ by summation over all relevant profiles:

$y_{s_1,s_{13}}(z,v) = \sum_j (p_1)^{a_{1j}} (1-p_1)^{b_{1j}} (p_2)^{a_{2j}} (1-p_2)^{b_{2j}} \cdots (p_5)^{a_{5j}} (1-p_5)^{b_{5j}} r_j$, where the sum is taken over all profiles that have v relapses and z surgical operations, and the suffix j is used to illustrate the values that are specific to each such profile. For the sake of effective computation, we traverse the whole set of profiles once, for each value of θ, and aggregate the probability of each profile into the corresponding $y_{s_1,s_{13}}(z,v)$, i. e. we compute these probabilities for all combinations of values of z and v, simultaneously. This method is used to compute the likelihood. The complete set of profiles can be prepared in advance, and this means that some of the time-consuming work is performed just once. In addition, since there are fewer profiles than unique pathways through the model, the amount of work that is performed during the maximum likelihood estimation is less using profiles.

3.3 Goodness of fit of the estimator

We used the estimated models, to simulate disease activity in each model for a period of ten years, for as many patients as we had in the input data. For each patient, each year was classified into disease course levels according to relapses and surgery experience. This predicted disease course was compared to the observed disease course from the 147 patients (Table 1). The predicted disease course should be similar to the observed disease course.

A validation of the estimator itself was performed using the above-described procedure backwards. Using known parameter values, we simulated the model and created a number of exercise dataset, with number of relapses and number of surgical operations per year for each individual. The Maximum Likelihood estimator was then used on these datasets to estimate the parameters used for simulation. The residuals between the estimated parameter values and the original (true) parameter values were computed, across all training datasets. This was done both for the probabilities in the model and for the mean length of stay in remission, and relapse, which are functions of the probabilities.

3.4 Uncertainty in estimates

We used bootstrap analysis to address the uncertainty in the parameter estimates [14]. The original dataset was replicated, by resampling the patients along with their corresponding disease history data. This yielded a set of replicated parameter vector estimates which was used to estimate variation and confidence intervals. To use the bootstrap results in a stochastic evaluation of the model, one could sample whole parameter vectors from the set of replicates, so that the dependencies between the parameters are kept intact [15].

3.5 Software tools.

The maximum likelihood estimator was implemented as a package in the R language and environment for statistical computing [16], with some components written in the C language [17]. The package can be obtained from the corresponding author. The bootstrap analysis used the boot package for R [18, 19].

4. Results

The estimated model parameters p_1, p_2, ..., p_5 for CD and UC are presented in Table 3. The probabilities of remission, p_1, p_2, are quite different, about twice as high during the first month (0.63) as during the following months (0.33), in CD. The difference is about seven times in UC (0.81 first month, and 0.12 during the following months). The different probabilities mean that the probability of remission declines over time, as expressed in time since start of relapse. The probabilities of relapse, p_3, p_4, are similar between CD and UC, with quite a large different between the first and following months (0.40 vs. 0.023 in CD, 0.49 vs 0.022 in UC), also here demonstrating a time-dependence, with probabilities that decline over time since start of remission. The probability of surgery, per month in relapse, p_5, is 0.19 in CD and 0.023 in UC.

The mean length of stay in remission in the model for CD, is 27.6 months, however the length of stays is skewed with a median of 9 months, which was estimated from simulation using the point estimates presented above. The mean length of a relapse is 2.1 months, and the median is one month. Using the first month of relapse (i.e. state 3 as shown in Figure 1), as a starting point, a patient is expected to suffer 1.8 relapses including the initial one (95% confidence interval (CI): 1 to 4 relapses), and spend a total of 2.7 months in relapse, and 9.3 months in remission (95% CI: 2 to 12 months), during a one-year time frame. The patient would receive

0.51 surgical operations during this time. The long-term rates of relapse and surgery, as the Markov model goes towards a steady state, are 0.48 relapses/year and 0.02 surgical operations/year.

In the model for UC, the mean length of stay in remission is 24.5 months, however the length of stays is very skewed with a median of 2 months. The mean length of a relapse is 2.7 months, and the median is one month. The 75% quantile of the relapse length of stay is one month too, so most stays in relapse last for just one month. With a starting point as above, a patient is expected to suffer 2.0 relapses including the initial one (95% CI: 1 to 5), and spend a total of 3.1 months in relapse, and 8.9 months in remission (95% CI: 0 to 12), during a one-year time frame. The expected number of surgical operations would be 0.072 during this period of time. The width of the 95% CI for total time in remission indicates that there is a very large variation between the patients in this model (c.f. Discussion). The long-term rates are 0.55 relapses/year and 0.004 surgical operations/year.

4.1 Goodness of fit of the estimation model

We used our estimated models, to simulate disease activity in each model for a period of ten years, aggregated this over individuals and calendar years, and compared it to the observed disease course (Figure 2).

For both diseases, there is an initial spike in *severe* and *mild*. Then, the disease course appears to go into a fairly steady state. In the observed disease course, the spike comes from patients getting the disease, and the steady state follows as the patients come under treatment. To model this, we use relapse as starting point when predicting the disease course.

For CD, the similarity is good. For UC the predicted disease course is still similar to the observed, but there is a decline in severe and mild in the observed disease course that does not appear in the predicted disease course (c.f. discussion).

We also performed a validation of the estimator itself by looking at residuals between estimated parameter values and the true parameter values when these were known in a number of exercise datasets of the same size as the CD dataset. For the probability estimates, the mean residual was -0.03 (SD 0.09). Using the probabilities to predict the mean length of stays in remission and relapse, the mean residual of the length of stay estimates was -0.4 (SD 1.6) months.

4.2 Uncertainty in estimates

The dataset for CD was bootstrapped to estimate standard deviation and confidence intervals for the parameters and for the duration of remission and relapse (Table 4). The bootstrap means of p_1 and p_2 are somewhat different than the original estimates (Table 3), but p_3, p_4, and p_5 are very similar. The SD:s of the probability estimates are in the range of 0.003 to 0.14. The bootstrap mean duration of remission is 29.9 months, and duration of relapse is 2.3 months (SD 9.5, 1.9, respectively), which is slighly higher than the original means. A 95% confidence interval for the mean duration of remission is 22 to 57 months (relapse 1.6 to 3.0 months).

5. Discussion

In this paper, we have presented a maximum likelihood estimator for a situation where patients have been observed indirectly, through disease history aggregated over yearly observation periods. We estimated an underlying Markov model by disaggregating the data,

and simultaneously rescaling time into a one month model cycle. The main part of our problem is to disaggregate our data, and as it is accomplished, the translation into shorter inteval comes automatically. Disaggregation and time translation is done by the same methodology. The translation from a long observation interval into a shorter perspective has been previously addressed [8,9,10,11,15]. Matrix decomposition only solves the part of our problem that is to translate into a shorter interval of time. The EM algorithm could in principle be employed to disaggregate our data, but where Craig and Sendi use the EM algorithm to impute values between their observations, we would need to use the algorithm to impute the entire disease history, for every patient. Another possibility would be to do a parametric or semiparametric formulation of the transition probabilities [11,15], and manipulate the probabilities directly to obtain a good fit to the data. However, like with the EM algorithm, having to rely on data on disease history aggregated over time, imposes a practical problem of disaggregation which we judge to be impractical using this approach.

We examined our maximum likelihood estimator using training datasets, which indicates that it works with data simulated from our particular estimation model. The model allows the probability of remission to depend on time since start of relapse, and the probability of relapse to depend on time since start of remission. In simulated training datasets with time-dependence embedded into the model structure in this way, the resulting estimates picks this up. Likewise, in training datasets without such dependence, the estimator usually gives estimates that do not show this dependence. This shows internal consistency of the estimator.

The estimator was optimized to improve execution time. We cannot isolate the time savings from the profiles method presented above, but the overall optimization of the estimator, where

the profiles provided a substantial part of the time savings, reduced the time to arrive at an estimate of about three hours down to about one minute, using a computer with a Intel Core Duo processor T2400 (1.83GHz) running Windows. The execution time applies to each one of the starting points for the numerical search, so the impact of optimization is considerable. For the whole work in total, the estimation took a couple of hours effective time, and the bootstrap analysis about eight hours. Craig & Sendi [8] demonstrate how another method, the E-M algorithm, can be used for difficult estimation problems. Since they do not give any indications on the performance of their estimator and since their approach applies only to a part of our problem, we cannot compare the execution times. They use the EM algorithm to impute values at a few time points between their observations, whereas our situation with aggregated data would require imputation of the entire process of every patient. Therefore the EM algorithm appears unwieldy for our situation with partial data aggregated over time. Craig and Fryback uses a MCMC approach to impute unobserved values. Their estimator used 48 CPU hours on a IBM RISC/6000 computer, after which it had produced a complete posterior distribution of the model parameters. This is a long time on a huge computer, but the full posterior distribution is also very useful and complete.

Early attempts with a very simple disease activity model, with just one relapse state and one remission state failed to fit to the given observations. Such a model has a very low probability of seeing the frequencies of silent, mild and severe patient-years such as in Table 1: they would instead be concentrated to just one of the three categories. This was the reason for splitting the relapse and the remission into a first month state and a state for the following months, respectively, allowing for the time dependence described above. Furthermore, the choice of model structure excluded death. UC does not appear to have any excess mortality, while CD has some [2]. We have assumed that the end of a patient's observation period was

non-informative for the disease activity, regardless of cause. It was our judgement that this would not affect the disease activity model if used with reasonably short time-frames.
Our exploration of the goodness of fit of the estimated models against the observed data, indicates that our estimator works fairly well also on the real dataset, at least for CD. Although our definition of disease course is simplified, it is used both for the predicted data and for the observed data so the validation is unaffected and we believe that the deviation from the clinically correct definition has a very minor impact. The bootstrap analysis resulted in fairly wide confidence intervals for some parameters, but this could be due to patient heterogeneity, and by the relatively little information available for the estimator. Silverstein et al [20] estimated a Markov chain for Crohn's disease by mapping patients into disease states over time according to detailed longitudinal medical records. They defined their disease severity states according to type of medical or surgical therapy and by the patients' response to medical therapy. In particular, relapse was split into four states (*drug-responsive*, *drug-dependent* and *drug-refractory*, and *surgery*), and remission was split into *remission*, and *post-surgery remission*. Therefore, a direct comparison cannot be made, since our disease states do not match theirs. However, we can make an approximate comparison to validate our model. They saw a median (75% percentile) stay in remission of 4 (19) months, 26 (76) months in post-surgery remission. Our remission state encompasses both, and our estimate of the median duration falls in between, 9 months (39 months). Silverstein estimated the median duration of relapse to 1.4 to 2.7 months, unless the patient was drug-dependent (7.7 months). Our estimated median was one month (75% percentile 2 months), so it was a little shorter. Silverstein et al used data that did not require any translation, and included a total of about 1,900 patient-years of observations compared to our 514, so they were less affected by censoring in states of long duration, and should have less uncertain estimates. This might contribute to why our estimate in remission is much lower that their estimate of the duration

in post-surgery remission. We believe that the remaining discrepancies are explained by the different disease state definitions.

Our disease activity model was designed primarily with CD in mind. UC is different since surgery methods exist that will end the disease activity. After colectomy, UC cannot relapse. However, pouchitises and IRA-proctitises can occur, so our relapsing and remitting model of UC could still be fairly valid as long as we define relapse as UC relapse, pouchitis and/or IRA-proctitis. An important question is whether the way UC relapse occurs before colecomy, and the way pouchitis/proctitis occurs after colectomy, are similar enough from a stochastic point of view. Our Markov model and the corresponding maximum likelihood estimator have time-homogeneous transition probabilities. Any change in the behaviour of the patients in connection to a colectomy would violate this underlying assumption. Indeed, we see a very surprising estimate of the duration of remission in UC, a mean of 24.5 months and a median of just two months. From a clinical point of view, such a short median duration is unrealistic. The distribution of the duration in the estimated model is skew with a very heavy tail, which explains the wide confidence interval for the time spent in remission. The goodness of fit test of predicted against observed disease course (Figure 2), shows an initial spike, and a fairly stable long-term distribution. Initially, when the disease is active and the patient is diagnosed, there are no patients in remission. This is handled by the estimator by using relapse as the starting state. In the long-term, Crohn's disease shows a fairly level distribution of *remission*, *mild* and *severe*, which we interpret as observing a steady state with some noise. Thus, we see no signs of violation of the time-homogeneous assumption in Crohn's disease during the observed time period. In ulcerative colitis, however, there is seemingly a growing proportion of patients in remission over time, i.e. a sign of time-inhomogeneous probabilities of relapse and remission. This is not visible in the predicted disease course, and should not be since the

disease activity model is time-homogeneous. Thus, we believe that ulcerative colitis due to its transient nature is not very well modelled by our disease activity model.

The predicted rates of relapse and surgery are initially high due to the starting point in relapse, and become lower in the long term. The observed average rates fall in between, which should be expected as they represent a mixture of the initial and long-term rates. We believe that the relation between the predicted and observed rates is reasonable.

In our disease activity model, there is no connection between surgery in relapse and a quick return to remission. Not capturing this dimension is a weakness of our model. Surgery could be embedded into the model structure to solve this problem, but that would make the maximum likelihood estimator more complicated, and we know from experience that the evaluation time of the estimator would be very long. Another aspect of surgery is that an UC patient could in real life have a colectomy with stomia during relapse, and elective pouch surgery in a subsequent period of remission, i.e. a two-session surgery [1,4]. This is entirely impossible in our model since we tie all surgery to relapse. This limitation to our model gives rise to additional doubt to whether it is suitable for UC. These aspects on surgery, and also the overall transient nature of UC through its curative surgery option would require a further developed disease activity model with it's own estimator. The increased complexity of such a model would make the estimator much slower, imposing a strong practical limitation. However, after succesful optimization, the impact would be considerably less and this could be attempted in a future study.

Our aim to develop a model in which a hypothetical intervention effect could be explored, appears to have been met, at least for Crohn's disease. We established a link from the

observed annually aggregated data to a model with one month cycles, which we could verify. This model represents the mix of treatment options in use during the observation period. A novel intervention could be compared to the standard care given by this mix, by modifying the model parameters. To explore the value of, say shortening relapses by some amount, one would modify the probabilities of remission accordingly. By assigning unit costs and appropriate effectiveness measures, e. g. QALY weights, to each state, the costs and QALYs could be estimated with or without this intervention, and a cost-effectiveness analysis be performed. The uncertainty surrounding the model parameter estimates could be incorporated based on the bootstrap analysis as outlined above.

Acknowledgements

The authors wish to thank Georg Lindgren and Tobias Rydén, Lund University, Sweden, for valuable advice regarding the maximum likelihood estimator.

Conclusions

The maximum likelihood estimator that translates yearly observations of aggregated disease history into monthly transition probabilities appears to work. The estimated disease activity model for Crohn's disease validates well to observed data and has good face validity. The disease activity model is less suitable for ulcerative colitis, due to its transient nature through the presence of curative surgery.

References

1. Nilsson Å, Tarmsjukdomar [Bowel Diseases], In: Asplund K, Berglund G, Lindgren S, Lindholm N (Eds), Internmedicin [Internal Medicine], Sweden: Liber; 2002.

2. Jess, T., Riis, L., Vind, I., et al, Changes in clinical characteristics, course and prognosis of inflammatory bowel disease during the last five decades: a population-based study from Copenhagen, Denmark, Inflamm Bowel Dis 2007;13:481– 489.

3. Munkholm P, Langholz E, Davidsen M, Binder V. Disease activity courses in a regional cohort of Crohn's disease patients.Scand J Gastroenterol. 1995 Jul;30(7):699-706.

4. Lindmark G, Ulcerös kolit och Crohnkolit [Ulcerative colitis and Crohn's disease], In: Jeppsson B, Peterson H-I, Risberg B (Eds), Kirurgi [Surgery], Sweden: Studentlitteratur; 2001.

5. Langholz E, Munkholm P, Davidsen M, Binder V. Course of ulcerative colitis: analysis of changes in disease activity over years.Gastroenterology. 1994 Jul;107(1):3-11.

6. Miller DK, Homan SM. Determining transition probabilities: confusions and suggestions. *Med Decis Making* 1994; 14: **52-58**.

7. Fleurence RL, Hollenbeak CS. Rates and probabilities in economic modelling: transformation, translation and appropriate application. Pharmacoeconomics 2007;25:3–6.

8. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain.Health Econ. 2002 Jan;11(1):33-42.

9. Welton NJ, and Ades AE (2005), Estimation of Markov Chain Transition Probabilities and Rates from Fully and Partially Observed Data: Uncertainty Propagation, Evidence Synthesis and Model Calibration., Medical Decision Making, 25, 633-645.

10. Charitos T, de Waal PR, van der Gaag LC, Computing short-interval transition matrices of a discrete-time Markov chain from partially observed data. Statistics in Medicine 27(6), 905-21.

11. Craig, B. A., Fryback, D. G., Klein, R., and Klein, B. E. K. (1999), "A Bayesian approach to modelling the natural history of a chronic condition from observations with intervention," Statistics in Medicine, 18(11), 1355-1371.

12. Lindgren, B. W. Statistical theory, 3rd ed. Collier MacMillan Publishers, London; 1976.

13. Ross, S. M., Introduction to probability models, 4th ed, Academic Press Ltd, London; 1989.

14. Hjorth JSU. Computer Intensive Statistical Methods, Validation, Model Selection and Bootstrap. London: Chapman and Hall, 1994.

15. Borg S, Ericsson Å, Wedzicha J, Gulsvik A, Lundbäck B, Donaldson GC & Sullivan SD. A Computer Simulation Model of the Natural History and Economic Impact of Chronic Obstructive Pulmonary Disease. Value Health. 2004 Mar-Apr;7(2):153-67.

16. R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

17. Kernighan, B. W., Ritchie, D. M., The C programming language, New Jersey, USA: Prentice-Hall; 1978.

18. Angelo Canty and Brian Ripley (2008). boot: Bootstrap R (S-Plus) Functions. R package version 1.2-32.

19. Davison, A. C. & Hinkley, D. V. (1997) Bootstrap Methods and Their Applications. Cambridge University Press, Cambridge. ISBN 0-521-57391-2.

20. Silverstein, M. D., Loftus, E. V., Sandborn, W. J., Clinical Course and Costs of Care for Crohn's Disease: Markov Model Analysis of a Population-Based Cohort, Gastroenterology 1999; 117, 49-57.

Figures and tables

Figure 1: The disease activity model with one-month cycles. The probabilities within braces denote the probability of surgery. The state numbers 1-4 are shown within the little circles.

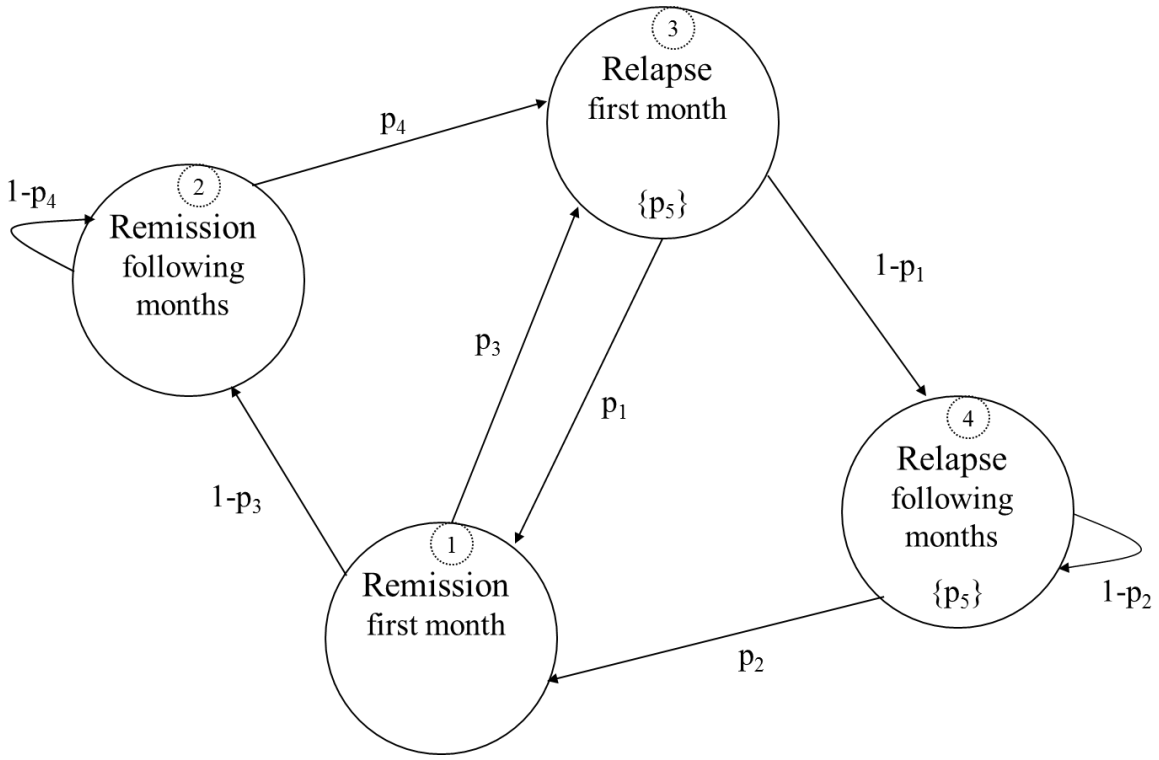


Table 1: The number of patients in the cohort with Crohn's disease (CD) or ulcerative colitis (UC), and the number of observed patient-years that they provided in each level of disease course[1], by diagnosis, and in total.

	Patients	Patient-years			
		Silent	Mild	Severe	Total
CD	58	318	110	86	514
UC	87	454	255	69	778

Notes

1. Each observed year is assigned a level of disease course: A year without surgery or relapses is a *silent* year. A relapse is defined as a contact with a physician due to gastrointestinal symptoms, that led to increased use of medication or surgery. A year with 1-2 relapses and no surgery is a *mild* year. A year with surgery and/or more than two relapses is a *severe* year. Some patients were not observed with all levels of disease course.

Table 2: The number of patient-years observed, split by number of relapses per year, and by number of surgical operations per year, respectively, in Crohn's disease (CD) and ulcerative colitis (UC).

		Number of relapses per patient-year							
		0	1	2	3	4	5	6	7
Relapses[1]	CD	318	123	44	22	2	5	0	0
	UC[2]	454	208	66	28	10	4	6	2
		Number of surgical operations per patient-year							
		0	1	2	3	4	5	6	7
Surgical operations	CD	438	50	14	5	5	2	0	0
	UC	753	18	6	0	1	0	0	0

Notes

1. A relapse is defined as a contact with a physician due to gastrointestinal symptoms, that led to increased use of medication or surgery.
2. Seven relapses/year is impossible in our model. These observations were recoded to 6 relapses.

Table 3: Estimates of the probability parameters and estimates of the durations of remission and relapse, in the disease activity models for Crohn's disease (CD), and ulcerative colitis (UC).

	Remission duration[1], months.		Relapse duration[2], months.		Probabilities[3]				
	Mean	(Median)	Mean	(Median)	p_1	p_2	p_3	p_4	p_5
CD	27.6	(9)	2.1	(1)	0.63	0.33	0.40	0.023	0.19
UC	24.5	(2)	2.7	(1)	0.81	0.12	0.49	0.022	0.023

Notes:

1. The duration of one period of remission.
2. The duration of one relapse.
3. These are probabilities of remission (p_1, p_2), probabilities of relapse (p_3, p_4), and probability of surgery in the relapse states (p_5) . See also Figure 1.

Table 4: Results from the bootstrap analysis[1] of the CD model parameters. Bootstrap mean, standard deviation (SD), and 95% confidence intervals (CI)[4].

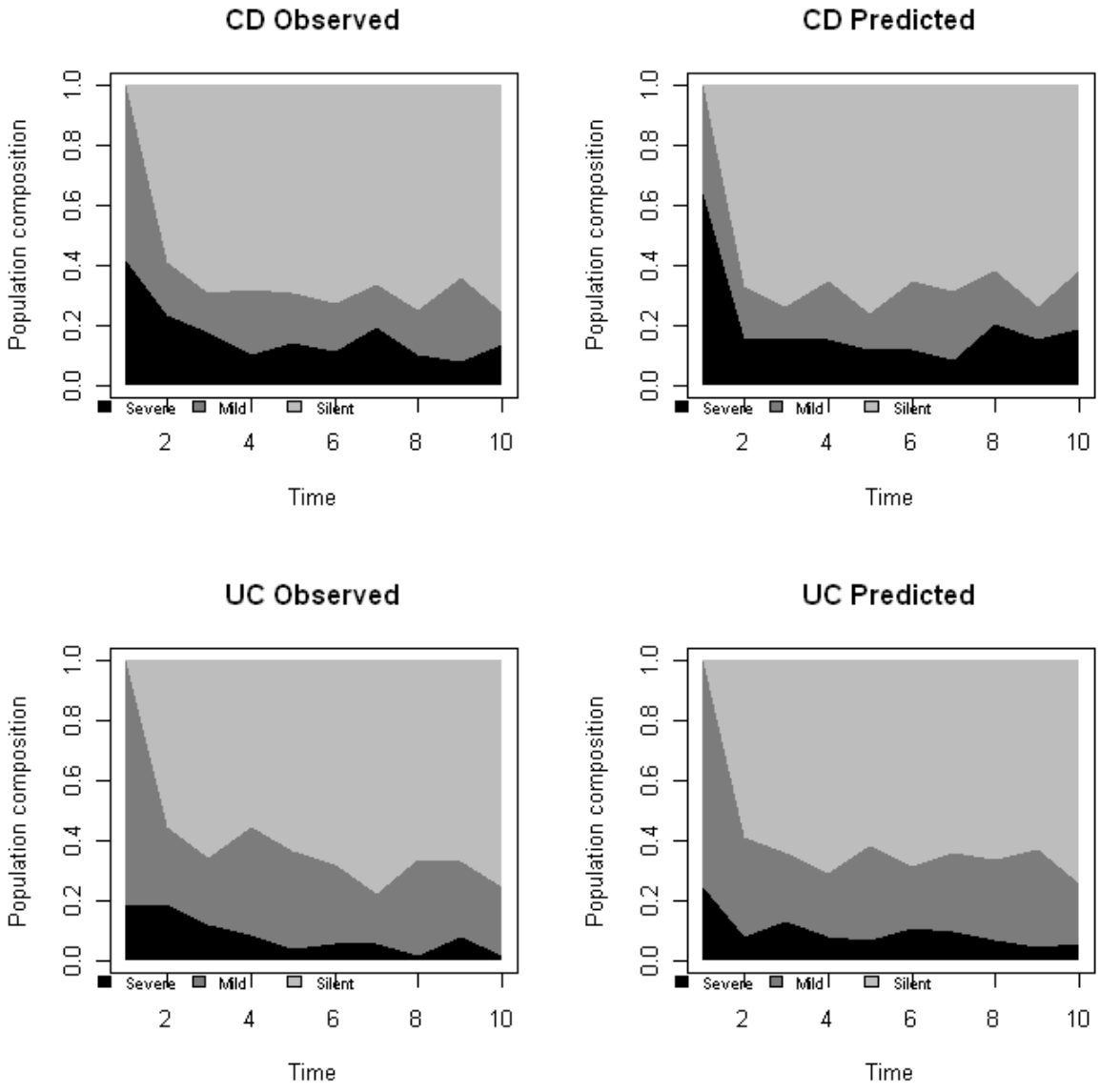
	p_1	p_2	p_3	p_4	p_5
Mean	0.58	0.36	0.39	0.022	0.19
SD	0.14	0.12	0.06	0.003	0.04
95% CI	(0.21, 0.80)	(0.19, 0.64)	(0.29, 0.48)	(0.01, 0.03)	(0.12, 0.27)

	Remission duration[2], months.	Relapse duration[3], months.
Mean	29.9	2.3
SD	9.5	1.9
95% CI	(22.3, 57.1)	(1.6, 3.0)

Notes:

1. The analysis used 500 bootstrap replicates.
2. The duration of one period of remission.
3. The duration of one relapse.
4. The confidence intervals have individual confidence levels of 95%. Simultaneous considerations of more than one interval results in a lower confidence level.

Figure 2: Disease course during ten years time, as observed in the cohort (Observed) and predicted using the estimated disease models (Predicted) , for CD and UC. During any given year (horisontal axis), the colours show the proportions of patients in each level of disease course[1]; *severe* (black), *mild* (dark grey) and *silent* (light grey).



Notes:

1. A year without surgery or relapses is a *silent* year. A relapse is defined as a contact with a physician due to gastrointestinal symptoms, that led to increased use of medication or

surgery. A year with 1-2 relapses and no surgery is a *mild* year. A year with surgery and/or more than two relapses is a *severe* year.