

# A new likelihood model for analyses of pharmacoepidemiologic case-control studies which avoids decision rules for determining latent exposure status

Henrik Støvring (✉ [stovring@ph.au.dk](mailto:stovring@ph.au.dk))

Aarhus University: Aarhus Universitet <https://orcid.org/0000-0002-5821-2351>

Anton Pottegård

Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark

Jesper Hallas

Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark

---

## Research article

**Keywords:** case-control study, reverse waiting time distribution, maximum likelihood, parametric modelling, pharmacoepidemiology

**Posted Date:** September 21st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-70902/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Medical Research Methodology on July 8th, 2021. See the published version at <https://doi.org/10.1186/s12874-021-01312-y>.

1 **Title:** A new likelihood model for analyses of pharmacoepidemiologic case-control studies which  
2 avoids decision rules for determining latent exposure status

3  
4  
5 **Authors:** Henrik Støvring (1,2), Anton Pottegård (2), Jesper Hallas (2)

6  
7 1: Biostatistics, Department of Public Health, Aarhus University, Aarhus, Denmark

8 2: Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern  
9 Denmark, Odense, Denmark

10

11

12 **Corresponding author:**

Henrik Støvring

email: [stovring@ph.au.dk](mailto:stovring@ph.au.dk)

Department of Public Health

Phone: +45 8716 7991

Aarhus University

Fax: +45 8716 7305

Bartholins Allé 2

8000 Aarhus C

Denmark

13

14

15 **Word count** excluding abstract, tables, figures and references: 4,979

16

1 **Abstract**

2 **Background:** Case-control studies based on pharmaco-epidemiological databases typically use  
3 decision rules to determine exposure status from information on dates of prescription redemptions,  
4 although this induces misclassification. The reverse Waiting Time Distribution has been suggested  
5 as a likelihood based model to estimate the latent exposure status, and we therefore suggest to  
6 extend this into a joint likelihood based model, which incorporates both the latent exposure status  
7 and the exposure-outcome association. This will achieve consistency and efficiency of the  
8 estimates, i.e. they are asymptotically unbiased and have optimal precision.

9 **Methods:** We established a joint likelihood for the observed case-control status and last  
10 prescription redemption before the index date. The likelihood combines the ordinary logistic  
11 regression likelihood and the reverse Waiting Time Distribution, and allows inclusion of covariates  
12 in both parts to adjust for observed confounders. We conducted a simulation study of the new  
13 model and standard models based on decision rules for exposure and the probability of being  
14 exposed, respectively, to assess the relative bias and variability of estimates. Lastly, we applied the  
15 models to a case-control study on use of nonsteroidal anti-inflammatory drugs and risk of upper-  
16 gastrointestinal bleeding.

17 **Results:** In simulation studies the new model had low relative bias (<1.4%) and largely retained  
18 nominal coverage probabilities (90.2% to 95.1% of nominal 95% confidence intervals), also when  
19 moderate misspecification was introduced. All standard methods generally had substantial bias (-  
20 21.1% to 17.0%) and low coverage probabilities (0.0% to 68.9%). The new method estimated the  
21 effect of nonsteroidal anti-inflammatory drugs on risk of with upper-gastrointestinal bleeding  
22 hospitalization to 2.52 (1.59 – 3.45), whereas the other methods had estimates ranging from 3.52  
23 (2.19 – 5.65) to 5.17 (2.40 – 11.11).

24 **Conclusions:** Unlike standard methods, the proposed model consistently gave unbiased estimates  
25 with adequate coverage probabilities. The model demonstrates the potential for the reverse Waiting

1 Time Distribution to be integrated with existing likelihood based analyses in  
2 pharmacoepidemiological studies.

3

4 Word count for abstract: 309 (350 allowed)

5

6 **Keywords:** case-control study; reverse waiting time distribution; maximum likelihood; parametric  
7 modelling; pharmacoepidemiology

## 1 **Background**

2 Pharmacoepidemiologic case-control studies are commonly based on routinely collected register  
3 data, as this is an efficient strategy for assessing the risk of adverse drug effects in real-world  
4 settings. Typically in such studies, exposure information is obtained from databases recording all  
5 individual prescription redemptions for a population and then linked with information on for  
6 example hospital admissions for each individual in the population. Since the actual treatment after  
7 redeeming a prescription is not recorded, such studies routinely rely on decision rules for defining  
8 drug exposure status at the index time for both cases and controls. In a recent study, we suggested to  
9 use the reverse Waiting Time Distribution (WTD) to estimate the probability of being exposed on  
10 the index date, and use this probability as exposure variable.(1) The objective was to better allow  
11 for the fact that the actual exposure status is not recorded, and that the analyst may therefore  
12 introduce misclassification bias when forcing a dichotomy upon exposure. The approach was  
13 intended to directly represent the analyst's uncertainty about actual exposure by using probability of  
14 being exposed as exposure covariate. We found that the approach improved statistical efficiency  
15 when studying the risk of upper gastrointestinal bleeding (UGIB) associated with use of a  
16 nonsteroidal anti-inflammatory drug (NSAID).(1) The suggested method consisted of two steps:  
17 First, a parametric reverse WTD model was fitted to the last observed NSAID prescription  
18 redemption among controls before their index date to estimate their probability of being users on  
19 the index date. Second, we used the predicted probability as covariate in a logistic regression with  
20 case-control status as outcome.

21

22 However, both the traditional decision rule based approach and using the probability of exposure as  
23 covariate are susceptible to misclassification bias. Also, they are not able to incorporate the  
24 uncertainty regarding the actual exposure into the subsequent analysis estimating the effect of being  
25 exposed with respect to the outcome of interest. This will bias uncertainty estimates. To overcome

1 these biases, we here propose to base estimation on a joint parametric likelihood for the last  
2 observed prescription redemption in the year preceding the index date, if any, together with the  
3 observed case-control status. From a theoretical point of view, formulation of a full likelihood  
4 model is attractive as it provides estimates based on a coherent and transparent model of what is  
5 actually observed, i.e. the case-control status and the time of last redemption before the index date.  
6 Since the observed data are conceptually linked together via the unobserved (i.e. latent) exposure  
7 status, we can obtain estimates for the exposure-outcome relation by maximizing the likelihood.  
8 Based on general likelihood theory, such estimates will have optimal large sample properties.(2)  
9 Note that when no prescription redemption is observed in the year before the index date, we can  
10 safely assume that the person is unexposed, and thus such persons can also be readily included in  
11 the analysis.

12

13 In this paper, we first establish the new joint likelihood, which is a synthesis of a reverse WTD  
14 model and the ordinary logistic regression of exposure status among cases and controls. In a  
15 simulation study we then compare the performance of the new joint likelihood model with classic  
16 decision-rule based analyses as well as the previous suggestion of a two-step model based on  
17 exposure probabilities, and we investigate the impact of misspecification. We finally apply the  
18 models to data on NSAID prescription redemptions and hospitalization with UGIB in a Danish  
19 case-control study.

## 20 **Methods**

21 As a starting point, let us assume that if a person is in continued treatment with a drug, then (s)he  
22 will redeem prescriptions following a renewal process with  $f(t)$  as the inter-arrival density (IAD)  
23 between subsequent redemptions. When this renewal process is intercepted at an index date,  $R$  is  
24 the time from the last renewal (redemption) before the index date until the index date (Figure 1).

1 We let  $X$  denote whether a person continues having prescription redemptions ( $X = 1$ ) or not ( $X =$   
2  $0$ ) after the last observed prescription redemption before the index date. Note that we have in  
3 previous papers shown that even if the assumption of redemptions following a renewal process is  
4 violated, then the backward recurrence time (time from last prescription redemption before the  
5 index date to the index date) can still be reasonably modelled as if originating from a renewal  
6 process.(3,4)

7  
8 [Figure 1 around here]

9  
10 Concerning the outcome and exposure in the case-control study, a person can either be a case,  
11  $Y = 1$ , or a control,  $Y = 0$ , at the index date  $t_0$ , and be exposed to treatment,  $Z = 1$ , or not,  $Z = 0$ .  
12 We will assume that the case-control status of a person depends on the true, but unobserved binary  
13 exposure status,  $Z$ , in the ordinary fashion of logistic regression, i.e.

$$P(Y = 1|Z = z) = \frac{\exp(\beta_0 + \beta_1 z)}{1 + \exp(\beta_0 + \beta_1 z)}$$

14 In this formulation,  $\exp(\beta_1)$  is the odds ratio of interest when studying the association between  
15 being exposed and becoming a case. To maintain mathematical feasibility, we will assume that if  
16 the true exposure status is known, then whether a person is a case or a control is unrelated to the  
17 distribution of time since last prescription redemption. In other words, we assume  $Y$  and  $R$  are  
18 independent given  $Z$ .

19  
20 Let us now consider the relation between redemptions and exposure status. When a person  
21 continues treatment after the index date with one or more subsequent prescription redemptions  
22 ( $X = 1$  at the index date), then the person is by definition exposed at the index date ( $Z = 1$ ).  
23 However, when the last prescription redemption observed before the index date is also the last

1 redemption in the treatment of an individual ( $X = 0$ ), such individuals can be either treated ( $Z =$   
2 1) or untreated ( $Z = 0$ ) on the index date (Figure 1). This is because such individuals will remain  
3 exposed for some time after their last prescription redemption before they enter a state without  
4 treatment. We will assume that their duration of exposure after the prescription redemption,  $T$ ,  
5 follows the ordinary IAD of treated patients continuing treatment, i.e. the probability density  
6 function of  $T$  is  $f(t)$ . Finally, we will assume that persons without a prescription redemption in an  
7 interval before the index date,  $(t_0 - \delta; t_0)$ , are unexposed. To accommodate this, we will let  $V$  be  
8 an indicator variable, which is 1 if an individual has at least one redemption observed in  $(t_0 - \delta; t_0)$   
9 and 0 otherwise. Figure 1 gives an overview of the variables introduced and how they are related to  
10 four different characteristic patient types.

11

12 We can now write the likelihood contributions for each patient type by conditioning on  $X, Z$ , and  $V$ .  
13 In the following, we use  $pdf(\cdot)$  as a generic probability density function for the relevant stochastic  
14 variables which are written as arguments. Patient type numbering refers to Figure 1.

- 15 • *Patient type 1*: Patients continuing treatment, i.e.  $X = 1$ , which implies that  $Z = 1, V = 1$ :

$$\begin{aligned} pdf(R = r, Y = y|X = 1) &= pdf(R = r|X = 1) \cdot pdf(Y = y|X = 1, R = r) \\ &= g(r) \frac{\exp(\beta_0 + \beta_1)^y}{1 + \exp(\beta_0 + \beta_1)} \end{aligned}$$

- 16 • *Patient type 2A and 2B*: Patients stopping treatment, i.e.  $X = 0, V = 1$ :

$$\begin{aligned} pdf(R = r, Y = y|X = 0) &= pdf(Y = y|Z = 0) \cdot P(Z = 0|R = r, X = 0) \cdot pdf(R = r|X = 0) \\ &\quad + pdf(Y = y|Z = 1) \cdot P(Z = 1|R = r, X = 0) \cdot pdf(R = r|X = 0) \\ &= \frac{1}{\delta} \left( \frac{\exp(\beta_0)^y}{1 + \exp(\beta_0)} P(Z = 0|R = r, X = 0) \right. \\ &\quad \left. + \frac{\exp(\beta_0 + \beta_1)^y}{1 + \exp(\beta_0 + \beta_1)} P(Z = 1|R = r, X = 0) \right) \end{aligned}$$

1 where  $\frac{1}{\delta}$  is the uniform density associated with the last observed prescription redemption  
 2 among patients stopping treatment. The uniform density corresponds to assuming a constant  
 3 stopping rate with respect to the index date among treated.  $P(Z = 1|R = r, X = 0)$  is the  
 4 probability of a person still being treated at the index date (patient type 2A), even though the  
 5 person will stop treatment before any subsequent redemption. This is given by the  
 6 probability of the treatment duration  $T$  exceeding the time from the redemption to the index  
 7 date,  $r$ , i.e.

$$P(Z = 1|R = r, X = 0) = 1 - F(r)$$

8 where  $F(\cdot)$  is the cumulative density function of  $T$ . The density of  $T$  is assumed to be the  
 9 IAD of users continuing treatment, and from the definition of the BRD we have the  
 10 following relation:

$$g(r) = \frac{1 - F(r)}{M}$$

11 where  $M$  is the mean prescription duration,  $E(T)$ . By definition this implies that the  
 12 probability of being untreated at the index date for a patient stopping treatment (patient type  
 13 2B) is given by

$$P(Z = 0|R = r, X = 0) = F(r)$$

14 • *Patient type 3*: Patients without an observed prescription redemption in the time window  
 15  $(t_0 - \delta; t_0)$ , i.e.  $V = 0$ , and who are therefore unexposed ( $Z = 0$ ):

$$pdf(Y = y|V = 0) = \frac{\exp(\beta_0)^y}{1 + \exp(\beta_0)}$$

16 To complete the likelihood, we sum the three components after multiplication with their  
 17 probabilities of occurring, i.e.  $P(X = 1, V = 1)$ ,  $P(X = 0, V = 1)$ , and  $P(V = 0)$ , respectively.

18 With obvious notation, we can parametrize the probabilities as follows:

$$P(X = 1, V = 1) = P(X = 1|V = 1)P(V = 1) = p_x p_v$$

$$P(X = 0, V = 1) = P(X = 0|V = 1)P(V = 1) = (1 - p_x)p_v$$

$$P(V = 0) = (1 - p_v)$$

1 In sum, the likelihood contribution for each individual becomes

$$l(p_v, p_x, \beta_0, \beta_1, \theta | r, y) = p_x p_v \cdot g_\theta(r) \frac{\exp(\beta_0 + \beta_1)^y}{1 + \exp(\beta_0 + \beta_1)}$$

$$+ (1 - p_x) p_v \frac{1}{\delta} \left( \frac{\exp(\beta_0)^y}{1 + \exp(\beta_0)} F_\theta(r) + \frac{\exp(\beta_0 + \beta_1)^y}{1 + \exp(\beta_0 + \beta_1)} (1 - F_\theta(r)) \right)$$

$$+ (1 - p_v) \cdot \frac{\exp(\beta_0)^y}{1 + \exp(\beta_0)}$$

2 The full likelihood for all individuals is the product of the individual likelihood contributions. We  
 3 call this the joint likelihood as it is based on the joint density function for both the last prescription  
 4 redemption before the index date,  $R$ , and the case-control status,  $Y$ . In the joint likelihood,  $\theta$   
 5 denotes the parameters which the IAD, and thus the BRD, depends upon and which must be  
 6 estimated. If we choose a Log-Normal distribution for  $T$ , then  $\theta = (\mu, \sigma)$ , and we have the  
 7 following expressions for  $g_\theta$  and  $F_\theta$ :

$$g_{\mu, \sigma}(r) = \frac{\Phi\left(-\frac{\log(r) - \mu}{\sigma}\right)}{\exp\left(\mu + \frac{1}{2}\sigma^2\right)}$$

$$F_{\mu, \sigma}(r) = \Phi\left(\frac{\log(r) - \mu}{\sigma}\right)$$

8 where  $\Phi$  is the standard normal cumulative distribution function.

9

10 Note, that whether a case patient continues having prescription redemptions or not after the index  
 11 date may be hypothetical in the sense that case status may prohibit any further treatment. The model  
 12 does, however, only rely on the distribution of  $R$  which is in principle unaffected by any termination  
 13 of treatment after  $Y = 1$  has been observed (the case becomes a case).

14

1 Analogous with previous implementations of parametric WTDs, we have logit-transformed the  
2 probability parameters,  $p_x$  and  $p_v$ , and log-transformed the parameter  $\sigma$  which is required to be  
3 positive. Instead of a Log-Normal distribution, other parametric distributions with support on the  
4 positive axis could be considered, for example the Weibull or Gamma distributions.

5  
6 We have previously shown how parameters of the parametric reverse WTD can be allowed to  
7 depend upon covariates in a regression-like setup,(5) and the above formulation allows completely  
8 analogous inclusion of covariates. In particular, confounders which would ordinarily be included in  
9 an logistic regression analysis of the case-control study can be included in an equation for  $\beta_0$ . Also  
10 note that different parameters may depend on different sets of covariates.

## 11 12 **Simulation study**

13 The aim of the simulation studies was to examine the bias and precision of estimates. In the  
14 simulation study of the finite sample properties, we generated data compatible with the model  
15 developed above. More specifically, we used the following steps to create each dataset:

- 16 1. Generate a dataset of a given size with the actual latent exposure to treatment and  
17 case/control status for each individual. The dataset must satisfy that there is a specific  
18 number of controls for each case (1 or 10), and such that the exposure status reflects a pre-  
19 specified odds-ratio value, here 3. Further, we specified the proportion of individuals  
20 expected to have a prescription redemption observed in the observation window of length  
21 one year (15% and 25%, i.e. values for the probability  $p_v$  defined above), and how many of  
22 these that are expected to continue treatment after the index date (40% and 80%, i.e. values  
23 for  $p_x$ ). The latter two proportions indicate the prevalence on the index date among  
24 individuals with a prescription redemption in the observation window. Based on these  
25 values it is possible to derive the probabilities of exposure among cases and controls,

1 respectively, from which the exposure status can be drawn for all individuals in the dataset  
2 (see Additional file 1, Section A1 for details). Also for each individual it is randomly  
3 determined who has a prescription redemption in the observation window, and such that it is  
4 compatible with the random exposure status – exposed individuals must have a redemption  
5 in the observation window, whereas only some unexposed individuals will have such a  
6 redemption.

7 2. For the individuals assigned to continue treatment, and thus be exposed on the index date,  
8 we draw a prescription redemption before the index time, which follows a Log-Normal  
9 BRD. The parameters of the BRD,  $(\mu, \sigma)$ , are chosen such that they correspond to IADs  
10 with a median duration of 1.5 or 2 months, i.e.  $\mu = \log(1.5/12)$  or  $\log(2/12)$ , and  $\sigma =$   
11  $\exp(-0.35)$  or  $\exp(-0.25)$ , respectively. The values for both parameters were similar in  
12 magnitude to estimates found in the analysis of the application presented below.

13 3. For individuals identified to stop treatment after index date, we randomly draw a uniformly  
14 distributed time of prescription redemption in the observation window, but such that for  
15 those still exposed on the index date their redemptions follow the above specified Log-  
16 Normal BRD. We used rejection sampling to satisfy the requirement of the marginal  
17 distribution being uniform among all treatment stoppers when redemptions of the exposed  
18 sub-population followed a Log-Normal BRD (see Additional file 1, Section A2 for details).

19  
20 To examine the impact of a misspecified BRD, we also generated data with a Weibull BRD and  
21 analyzed the data with a model based on a Log-Normal BRD. To make the settings comparable, we  
22 used parameters such that the Weibull IAD had the same mean and variance as the Log-Normal  
23 IAD corresponding to the  $\mu$  and  $\sigma$  values stated above.

1 We generated datasets which consisted of 19,800 or 39,600 individuals as this either resembled the  
2 application below or had half the size of this application. For each scenario we generated 2,500  
3 datasets, which we analyzed with five different approaches:

- 4 1. Logistic regression where exposure is the actually assigned exposure status, i.e. exposure is  
5 not latent and thus there is no misclassification bias.
- 6 2. The proposed joint likelihood model for case/control status and the rWTD, as developed  
7 above
- 8 3. Logistic regression with the estimated probability of being in treatment as exposure variable.  
9 The probability is estimated from an ordinary rWTD analysis.
- 10 4. Logistic regression where exposure status is based on a 90 days window – individuals are  
11 considered exposed if they have a prescription redemption within 90 days prior to their  
12 index date.
- 13 5. Logistic regression where exposure status is based on a 30 days window – individuals are  
14 considered exposed if they have a prescription redemption within 30 days prior to their  
15 index date.

16 All logistic regression analyses used case-control status as outcome. The first analysis is the  
17 reference analysis, as this is the analysis that would have been used for analyzing a case-control  
18 study with directly observed exposure status. It can be expected to yield unbiased estimates and  
19 have optimal statistical efficiency (smallest uncertainty intervals). We therefore benchmarked the  
20 other four analyses against it. Our primary interest was to establish whether the new joint model  
21 achieved its theoretical, asymptotic superiority in samples with a realistic size. For the two last  
22 approaches, we kept all generated individuals in the analysis to maintain comparability of standard  
23 errors with the other analyses, even though it is common practice in pharmacoepidemiologic studies  
24 to exclude individuals that are not classified as non-exposed (i.e. used as reference) or exposed (i.e.  
25 have a redemption inside the prespecified main exposure window).(1,6)

1

2 For each method and setting we estimated the median relative bias of the estimated OR, the median  
3 standard error of the log-OR, the coverage probability of nominal 95% confidence intervals and  
4 finally the relative variance inflation factor with respect to the optimal analysis when exposure  
5 status is directly observed.

6

## 7 **Application**

8

### 9 **Cases and controls**

10 As case material we used a case-control study also employed in previous publications, which was  
11 described in detail there.(1,6) Briefly, our source population was the residents of Funen County,  
12 Denmark, during a study period of August 1, 1995 to July 31, 2006. We included as cases all  
13 subjects who were admitted to a hospital on Funen with UGIB, validated by single case review.  
14 Controls were selected by a risk-set sampling strategy, i.e. for each case, we randomly selected ten  
15 controls among the subjects in our source population who matched the case by sex and birth year.  
16 Controls were assigned an index date identical to the outcome defining date of the corresponding  
17 case. We required that both cases and controls had been residents of Funen for at least one year on  
18 the index date. As some of the very old cases had less than 10 eligible controls, the final control to  
19 case ratio deviated slightly from 10:1.

20

### 21 **Data analysis**

22 Although data was collected with matching of controls to cases, we did not analyze the data  
23 conditional on match sets. Each matched set consisted of at most 11 individuals, which did not  
24 contain sufficient information to reliably estimate parameters of a WTD for each matched set.  
25 Consequently, we ignored the matching and instead included the matching variables age and sex as

1 covariates in the association between exposure and case-control status, although this will likely  
2 introduce a small bias towards the null.(7)

3

4 As reference analyses we considered logistic regression with case-control status as outcome and the  
5 exposure covariate with respect to NSAID defined as follows:

- 6 • WTD probability – the probability of being exposed as estimated by a reverse waiting time  
7 distribution among controls in the year prior to their index date. We let the WTD parameters  
8 depend on age, sex, quantity dispensed (in DDD), choice of ibuprofene (the dominant  
9 NSAID), concurrent use of proton pump inhibitors, a diagnosis of rheumatoid arthritis,  
10 psoriasis arthritis or spondylarthritis, and concurrent use of methotrexate or systemic  
11 corticosteroids. The latter variables (use of proton pump inhibitors and onwards) were  
12 intended to reflect markers of long-term NSAID use. Based on estimated parameters,  
13 observed covariates and date of last NSAID dispensing before the index date, we calculated  
14 the probability of being exposed on their index date for each case and control. For further  
15 details, see our previous paper where this was termed the *full multivariable model*.(1)
- 16 • 90 days – exposed if at least one NSAID redemption was observed within 90 days before the  
17 index date
- 18 • 30 days – exposed if at least one NSAID redemption was observed within 30 days before the  
19 index date

20 For all the above models, we estimated a crude OR without adjustment for other variables and an  
21 adjusted OR. For the adjusted ORs, the following potential confounders were included as covariates  
22 in the logistic regression: 1) current use of the following drugs: vitamin K antagonists (VKA),  
23 aspirin, other antiplatelet drugs, dipyridamol, beta-blockers, selective serotonin reuptake inhibitors  
24 (SSRIs), systemic corticosteroids, proton pump inhibitors (PPIs), H2 receptor antagonists, statins,  
25 nitrates, spironolactone, calcium antagonists, bisphosphonates; 2) any history of the following

1 events: NVUGB, *Helicobacter pylori* (HP) eradication, peptic ulcer, chronic obstructive pulmonary  
2 disease (COPD), diabetes, ischemic heart disease, heart failure, stroke, hypertension, inflammatory  
3 bowel disease, malignant disease, renal failure; and 3) prescription or diagnosis markers of smoking  
4 or excessive alcohol consumption. For all drugs used as covariates, current drug use was defined by  
5 the redeeming of a prescription within less than 120 days before the index date.

6  
7 For the joint likelihood we let the parameters associated with the reverse WTD ( $\mu$ ,  $\sigma$  and  $p_x$ ) depend  
8 on the same covariates as the reverse WTD described above when using probability of exposure as  
9 covariate, i.e. the following variables: age, sex, quantity dispensed (in DDD), use of ibuprofene (the  
10 dominant NSAID), concurrent use of proton pump inhibitors, a diagnosis of rheumatoid arthritis,  
11 psoriasis arthritis or spondylarthritis, and concurrent use of methotrexate or systemic  
12 corticosteroids. For the logistic regression part of the joint likelihood we either omitted or included  
13 the same covariates (for  $\beta_0$ ) as described above for the adjusted models. Correspondingly, we  
14 labeled the two models either crude or adjusted, although both included adjustment in the reverse  
15 WTD part of the model.

16  
17 All statistical analyses were conducted in Stata 15.(8) A dedicated software package (ccwtdttt)  
18 implementing the method is provided at the IDEAS repository (<http://ideas.repec.org>) and may be  
19 installed in Stata using a search for the package name, i.e. `–search wtdttt, all–`.

## 20 **Results**

### 21 *Simulation study*

22 Results of the simulation study are shown in Table 1 for the setting with a sample size of 39,600; a  
23 fraction of 80% of patients continuing treatment; 25% of all individuals having at least one  
24 prescription redemption in the time window of one year prior to the index date; and an OR of 3.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

The new method based on a joint likelihood for the case status and last observed prescription redemption before the index date was virtually unbiased (range 0.3% to 1.7%) with coverage probabilities close to the nominal value (range 90.2% to 95.1%) and relatively low VIFs (range 1.09 to 1.24). With 1:1 matching and a low  $\mu$  value (median prescription duration of 1.5 months) there were a small number of datasets for which the likelihood estimation did not converge (< 70 of 2,500).

[Table 1 around here]

Results were largely similar for the joint likelihood method when the sample size was halved to 19,800, although with higher coverage probabilities (see Additional file 1, Table A1) or when the fraction of patients continuing treatment was reduced to 40% (see Additional file 1, Table A2). VIFs were however increased when fewer continued treatment (range 1.23 to 1.53). The other methods remained biased with low coverage probabilities, although their internal rank of performance changed in some settings.

In all settings there was substantial bias associated with the method of choosing a fixed exposure window, be it 30 or 90 days (relative bias ranged from -6.7% to -21.5%). Correspondingly, the coverage probabilities of nominal 95% confidence intervals for the odds ratio were low (range 0.3% to 40%). The variance inflation factor (VIF) varied considerably relative to the optimal analysis in which the true exposure status of individuals was directly observed (range 1.06 to 2.48), but given the substantial bias this is less relevant.

1 The analysis based on estimating the probability of exposure with the rWTD and then use this  
2 probability as exposure covariate also led to substantial relative bias (range 6.1% to 17.1%), which  
3 indicates that this approach in general overestimated the association. This approach also had low  
4 coverage probabilities (range 0.0% to 68.9%), but consistently it had high VIFs relative to the  
5 optimal analysis (range 1.56 to 2.00). The low coverage probabilities were thus not so much a  
6 consequence of small estimated standard errors, but rather of the magnitude of the bias.

7  
8 When the BRD was misspecified (data generated from a Weibull BRD and analyzed with a model  
9 based on a Log-Normal BRD), the relative bias increased slightly for the joint likelihood ( $\leq 1.7\%$ ),  
10 coverage probabilities remained near the nominal level ( $\geq 90.1\%$ ), and VIFs were as above ( $\leq$   
11 1.25). Similar results were seen with the smaller sample size of 19,800 (see Additional file 1, Table  
12 A3). The bias, coverage and VIFs for the other analytic approaches were similar to the settings  
13 where a Log-Normal BRD was used for generating the data.

### 15 *Application*

16 The characteristics of cases and controls are shown in an additional table (see Additional file 1,  
17 Table A4). The joint likelihood model provided an unadjusted estimate of the NSAID effect on risk  
18 of UGIB similar to the other models (OR = 5.57 (5.08 - 6.05)), although the estimate based on the  
19 probability of being exposed was higher (Table 3). However, when the association was adjusted for  
20 confounders, the estimated effect (2.52 (1.59 - 3.45)) was substantially lower for the joint model  
21 than estimates from the three other models. The joint model had better precision than the methods  
22 based on a fixed exposure window, but lower than the exposure probability approach. As in the  
23 previous papers, when reporting estimates based on a fixed exposure window we omitted patients  
24 who had an NSAID prescription outside the exposure window resulting in a lower sample size for  
25 these two analyses (30 days: n=9,453 and 90 days: n=12,662 vs the full sample size of 39,119). The

1 gain in precision for the models based on the WTD is thus due to their ability to include the entire  
2 sample in the estimation. Estimates were largely unchanged, whether we used conditional logistic  
3 regression to account for the age- and sex-matched design or included sex and age as covariates in  
4 ordinary logistic regression with all three standard analyses. However, for the adjusted analyses,  
5 estimated ORs were smaller for the two analyses based on a fixed exposure window (30 days or 90  
6 days).

7

8 [Table 3 around here]

## 9 **Discussion**

10 We have shown how it is possible to establish a likelihood for the effect of a latent exposure status  
11 in a case-control study, where the latent exposure status is modelled via a reverse Waiting Time  
12 Distribution (WTD). The reverse WTD relies on data regarding the last prescription redemption  
13 before the index date, if any, which are observed in pharmacoepidemiologic databases. Simulation  
14 studies showed that the developed model provided unbiased estimates with valid confidence  
15 intervals in settings with realistic sample sizes, even with a misspecified model. This was not the  
16 case for the two standard approaches with fixed exposure windows, nor the method based on using  
17 estimated exposure probability as covariate. The precision of the joint likelihood estimates was also  
18 better than the other methods as it had Variance Inflation Factors smaller than 1.25 relative to a  
19 model based on observing the true exposure status of all individuals. In the application, the joint  
20 likelihood model gave a lower adjusted estimate of the risk of UGIB associated with NSAID use  
21 than other methods. The joint likelihood model can be estimated with a Stata package we have  
22 made available online.

23

1 Theoretically, a full likelihood-based model provides the optimal analysis in terms of statistical  
2 efficiency, since it for large samples provides unbiased estimates with minimal standard errors  
3 (Cramér-Rao lower bound).(2) Other methods may yield smaller standard errors, but are then not  
4 consistent as they are asymptotically biased. We therefore joined the likelihood of the reverse WTD  
5 and that of a logistic regression to allow direct estimation of the OR in case-control studies with  
6 exposure information based on registry data. We based our approach on noting that logistic  
7 regression in a case-control study models the relation between exposure and case-control status.  
8 When using pharmacoepidemiologic register data, exposure status is not directly observed (did the  
9 patient actually take the pill on the index date?), but this latent exposure status is modelled in the  
10 reverse WTD. The reverse WTD links information on last observed prescription redemption before  
11 the index date to two different sub-groups of patients, those who are continuing treatment on the  
12 index date and those who are stopping treatment. In our approach we allowed for the fact that  
13 patients, who have no dispensings after their last observed prescription redemption prior to the  
14 index date, may still be in treatment on that index date, as they use up the dispensed quantity. As  
15 expected, the model provided virtually unbiased estimates for realistic sample sizes with valid  
16 confidence intervals when analyzing data generated from the same model as that used for analysis,  
17 i.e. when the model is correctly specified. Somewhat surprising for us, this was not the case for the  
18 method based on probability of exposure, but this is due to misspecification of exposure status: If  
19 exposure status is truly binary, as is the case in our simulation, then any estimated probability of  
20 exposure represents a misclassification of the individual's exposure status, which will result in  
21 bias.(10)

22

23 We are not aware of other methods that allow estimation of the exposure OR in a case-control study  
24 in which exposure status is latent and linked to observed prescription redemptions. Abrahamowicz  
25 and co-authors investigated how different models for the exposure-outcome relation over time could

1 be used in prospective studies on the risk of adverse effects with pharmacoepidemiological registry  
2 data.(11) They relied on model diagnostic criteria to identify the correct model among many  
3 candidates and found substantial bias associated with misspecification of the exposure-outcome  
4 relation with respect to timing. In our model we assumed that it was exposure on the index date that  
5 induced a potential effect on case-control status.

6  
7 While theoretically superior, our developed method has several limitations. First, the method does  
8 not allow taking into account any matching. This is expected to dilute effect estimates.(12)  
9 Including matching variables as covariates in the estimation may however remedy this, at least  
10 partially, and we did indeed see that this strategy provided similar results with the standard analyses  
11 in our application. Second, even when the model is correctly specified, it is complex and the  
12 maximum likelihood estimation procedure may occasionally not converge. Third, the method relies  
13 on the simplifying assumptions of the BRD being similar between cases and controls, and that  
14 patients having no dispensings after the index date may remain in treatment for a period of time  
15 after their last prescription redemption with a distribution of prescription durations similar to that  
16 seen among patients continuing treatment. Fourth, while we did examine the method's performance  
17 with some misspecification of the BRD, it is not clear how vulnerable the method is to more severe  
18 model-misspecification. For mathematical feasibility we modeled event times and case-control  
19 status as independent given the latent exposure status, but this may not hold for NSAID and UGIB,  
20 where we expect recent initiators to have a higher risk than later in the treatment phase.(13)

21 Addressing this will require further modelling, but we have likely underestimated the effect of  
22 NSAID on UGIB in our analysis based on the developed joint likelihood. Finally, the developed  
23 model only estimates the effect of current use, i.e. the effect of being exposed on the index date.

24

## 1 **Conclusions**

2 We have developed a novel likelihood-based method which provides direct estimation of the  
3 exposure effect in case-control studies with exposure information obtained from  
4 pharmacoepidemiological data. The most important contribution is to show how formal extension  
5 of the reverse WTD model allows flexible estimation without use of expert input on expected  
6 patterns of drug use for definition of exposure status. Specifically, the established model provides  
7 estimates relying only on the actually observed data, i.e. case-control status and last prescription  
8 redemption before the index date, possibly supplemented with information on confounders that can  
9 be seamlessly included in the estimation. The set of covariates included can potentially vary  
10 between parameters in the joint likelihood, such that for example the reverse WTD part of the  
11 model can rely on covariates informative for drug use patterns, while the effect parameter can be  
12 adjusted for confounding variables. The method can be expected to provide a similar precision to a  
13 study with no exposure misclassification, if it has a 25 – 50% larger sample size. The developed  
14 model only applies to case-control studies, but may provide a starting point for development of  
15 models for other study designs.

16

17

## 18 **List of Abbreviations**

19 BRD: Backward recurrence density

20 IAD: Inter-arrival density

21 NSAID: Non-steroidal anti-inflammatory drugs

22 OR: Odds ratio

23 UGIB: Upper gastrointestinal bleeding

24 VIF: Variance Inflation Factor

25 WTD: Waiting Time Distribution

1 **Declarations**

2

3 **Ethics approval and consent to participate**

4 Approval from an Ethics Committee was not required according to Danish law (Law 502, May 23,  
5 2018, §10). The law implies that it is not required to obtain consent to participation for studies  
6 based on register data in Denmark. The data used in this study is an irreversibly anonymized  
7 version of a data set established in 2006, with permission granted from The Danish Data Board  
8 (“Datatilsynet”) protocol # 2006-41-7019. For an overview of the Danish registers and their use, see  
9 (9).

10

11 **Consent for publication**

12 Not applicable.

13

14 **Availability of Data and Materials**

15 Due to Danish laws on personal data, data cannot be shared publicly. To request these data,  
16 please contact the Danish Health Data Authority.

17

18 **Competing interests**

19 None relevant.

20

21 **Funding**

22 None, except for the public institutions where authors are employed. Institutions had no role in  
23 the design of the study and collection, analysis, and interpretation of data and in writing the  
24 manuscript.

1

2 **Authors contributions**

3 HS had the original idea for the study, derived the joint likelihood, conducted the simulations,  
4 analyzed the data and drafted a first version of the manuscript. JH provided access to data. AP and  
5 JH gave critical input to the design of simulations and the presentation of results, and critically  
6 reviewed the manuscript. All authors read and approved the final manuscript.

7

8 **Acknowledgement**

9 Not applicable.

## 1   **References**

- 2   1.    Hallas J, Pottegård A, Støvring H. Using probability of drug use as independent variable in a  
3       register-based pharmacoepidemiological cause-effect study—An application of the reverse  
4       waiting time distribution. *Pharmacoepidemiol Drug Saf.* 2017;26(12):1520–6.
- 5   2.    Silvey SD (Samuel D. *Statistical inference.* Chapman and Hall; 1975. 192 p.
- 6   3.    Støvring H, Vach W. Estimation of prevalence and incidence based on occurrence of health-  
7       related events. *Stat Med.* 2005 Sep 30;24(20):3139–54.
- 8   4.    Støvring H, Pottegård A, Hallas J. Determining prescription durations based on the  
9       parametric waiting time distribution. *Pharmacoepidemiol Drug Saf.* 2016 Dec;25(12):1451–  
10      9.
- 11  5.    Støvring H, Pottegård A, Hallas J. Refining estimates of prescription durations by using  
12      observed covariates in pharmacoepidemiological databases: an application of the reverse  
13      waiting time distribution. *Pharmacoepidemiol Drug Saf.* 2017;26(8):900–8.
- 14  6.    Dall M, Schaffalitzky de Muckadell OB, Lassen AT, Hansen JM, Hallas J. An association  
15      between selective serotonin reuptake inhibitor use and serious upper gastrointestinal  
16      bleeding. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc.* 2009  
17      Dec;7(12):1314–21.
- 18  7.    International Agency for Research on Cancer., Breslow NE, Day NE (Nicholas E. *Statistical*  
19      *Methods in Cancer Research. Volume 1, The analysis of case-control studies.* International  
20      Agency for Research on Cancer; 1980. 350 p.
- 21  8.    StataCorp. *Stata Statistical Software: Release 15.* College Station, TX: StataCorp LLC; 2017.
- 22  9.    Thygesen LC, Daasnes C, Thaulow I, Bronnum-Hansen H. Introduction to Danish  
23      (nationwide) registers on health and social issues: Structure, access, legislation, and  
24      archiving. *Scand J Public Health.* 2011;39(7 Suppl):12–6.

- 1 10. Satten GA, Kupper LL. Inferences About Exposure-Disease Associations Using Probability-of-  
2 Exposure Information. *J Am Stat Assoc.* 1993 Mar;88(421):200.
- 3 11. Abrahamowicz M, Beauchamp M-E, Sylvestre M-P. Comparison of alternative models for  
4 linking drug exposure with adverse effects. *Stat Med.* 2012 May 20;31(11-12):1014-30.
- 5 12. Breslow NE, Day NE, Halvorsen KT, Prentice RL, Sabai C. Estimation of multiple relative risk  
6 functions in matched case-control studies. *Am J Epidemiol.* 1978 Oct;108(4):299-307.
- 7 13. Moride Y, Abenhaim L, Yola M, Lucein A. Evidence of the depletion of susceptibles effect in  
8 non-experimental pharmacoepidemiologic research. *J Clin Epidemiol.* 1994 Jul;47(7):731-7.  
9  
10

1 **Figures**

2

3 Figure 1. Four typical persons (horizontal lines) representing the different types of contributions to the  
4 likelihood with respect to exposure status on the index date indicated by squares (filled: treated, unfilled:  
5 untreated). Black bullets represent prescription redemptions included in the likelihood, grey bullets are  
6 redemptions not included in the likelihood. The thicker black lines represent periods of treatment, whereas  
7 dashed lines are periods without treatment.  $T$  is duration of the prescription and  $R$  is time from the last  
8 prescription redemption before the index date  $t_0$  until the index date.  $X$  indicates whether a patient  
9 continues treatment,  $Z$  exposure status on the index date, and  $V$  whether a prescription is redeemed in the  
10 interval  $(\delta; t_0)$ . For simplicity, index dates have been aligned on the time scale. Type refers to patient type,  
11 see text for details.

1 Table 1. Simulation results, data generated with a Log-Normal Backward Recurrence Density.

		$\sigma=\exp(-0.35)$				$\sigma=\exp(-0.25)$			
	Analysis method (*)	Relative bias (%)	SE log(OR)	Coverage 95% CI (%)	VIF	Relative bias (%)	SE log(OR)	Coverage 95% CI (%)	VIF
$\mu=\log(1.5/12)$	<i>1:1 CC</i>								
	<b>True expo</b>	0.0	0.027	95.6	1.00	0.0	0.027	95.3	1.00
	<b>CC WTD</b>	0.4	0.029	95.1	1.15	0.8	0.029	93.5	1.19
	<b>WTD prob</b>	17.0	0.037	0.0	1.91	17.1	0.038	0.0	2.00
	<b>90 days</b>	-6.7	0.028	24.3	1.08	-7.8	0.028	14.2	1.12
	<b>30 days</b>	-12.1	0.037	5.4	1.90	-12.9	0.038	3.8	2.01
	<i>1:10 CC</i>								
	<b>True expo</b>	0.0	0.036	95.1	1.00	0.0	0.036	94.5	1.00
	<b>CC WTD</b>	0.3	0.038	95.0	1.09	0.7	0.038	94.2	1.11
	<b>WTD prob</b>	8.7	0.046	46.4	1.61	8.3	0.047	51.4	1.66
	<b>90 days</b>	-7.5	0.038	40.0	1.06	-9.1	0.038	25.6	1.08
	<b>30 days</b>	-18.2	0.046	1.0	1.60	-19.2	0.047	0.7	1.69
$\mu=\log(2/12)$	<i>1:1 CC</i>								
	<b>True expo</b>	0.0	0.027	94.4	1.00	0.0	0.027	94.4	1.00
	<b>CC WTD</b>	1.0	0.029	93.6	1.19	1.7	0.030	90.2	1.24
	<b>WTD prob</b>	14.4	0.036	0.8	1.84	14.1	0.037	1.3	1.91
	<b>90 days</b>	-8.1	0.029	14.1	1.18	-9.2	0.030	7.6	1.23
	<b>30 days</b>	-13.9	0.041	3.8	2.35	-14.5	0.042	3.9	2.48
	<i>1:10 CC</i>								
	<b>True expo</b>	0.0	0.036	94.8	1.00	0.1	0.036	94.7	1.00
	<b>CC WTD</b>	0.8	0.038	94.6	1.11	1.4	0.039	93.6	1.14
	<b>WTD prob</b>	6.7	0.045	62.6	1.56	6.1	0.046	68.9	1.60
	<b>90 days</b>	-10.2	0.038	17.2	1.12	-11.7	0.039	8.2	1.15
	<b>30 days</b>	-21.1	0.051	0.5	1.94	-21.5	0.052	0.3	2.03

2

3 The datasets had a sample size of 39,600, and on average 80% of patients continue treatment at the  
 4 index date, 25% of patients have a prescription redemption in the year before the index date and the  
 5 true OR is 3. For each setting 2,500 datasets were generated and analyzed.  $\mu$  and  $\sigma$  are parameters

1 of the assumed Log-Normal Backward Recurrence Density used for generating data, see text for  
2 details  
3  
4 (*\*) Analysis methods:*  
5 *1:1 CC* indicates 1 control per case, *1:10 CC* indicates 10 controls per case.  
6 **True expo** – logistic regression with the actual exposure status as covariate (the reference analysis).  
7 **CC WTD** – estimation based on joint likelihood for case-control status and the reverse WTD, Log-  
8 Normal Backward Recurrence Density.  
9 **WTD prob** – a reverse WTD with Log-Normal Backward Recurrence Density is estimated to  
10 predict the probability of an individual being exposed and this exposure probability is used as  
11 covariate in logistic regression.  
12 **90 days** – individuals are considered exposed if they have a redemption < 90 days before index  
13 date.  
14 **30 days** – individuals are considered exposed if they have a redemption < 30 days before index  
15 date, logistic regression.  
16

1 Table 2. Simulation results, data generated with a Weibull Backward Recurrence Density

		$\sigma=\exp(-0.35)$				$\sigma=\exp(-0.25)$			
	Analysis method (*)	Relative bias (%)	SE log(OR)	Coverage 95% CI (%)	VIF	Relative bias (%)	SE log(OR)	Coverage 95% CI (%)	VIF
$\mu=\log(1.5/12)$	<i>1:1 CC</i>								
	<b>True expo</b>	0.0	0.027	95.7	1.00	0.0	0.027	95.8	1.00
	<b>CC WTD</b>	0.9	0.029	94.2	1.15	0.5	0.029	95.1	1.18
	<b>WTD prob</b>	22.6	0.040	0.0	2.24	18.0	0.038	0.1	2.05
	<b>90 days</b>	-6.0	0.027	33.2	1.06	-8.1	0.028	11.2	1.13
	<b>30 days</b>	-11.1	0.035	6.7	1.69	-13.3	0.038	2.4	2.02
	<i>1:10 CC</i>								
	<b>True expo</b>	0.0	0.036	95.4	1.00	0.0	0.036	95.2	1.00
	<b>CC WTD</b>	0.6	0.038	94.9	1.08	0.5	0.038	95.2	1.10
	<b>WTD prob</b>	11.2	0.049	28.7	1.78	8.6	0.047	48.5	1.68
	<b>90 days</b>	-6.5	0.037	52.2	1.04	-9.5	0.038	20.6	1.09
	<b>30 days</b>	-16.8	0.044	1.1	1.45	-19.6	0.048	0.3	1.70
$\mu=\log(2/12)$	<i>1:1 CC</i>								
	<b>True expo</b>	-0.1	0.027	94.2	1.00	0.0	0.027	95.8	1.00
	<b>CC WTD</b>	1.7	0.029	90.1	1.21	1.6	0.030	91.5	1.25
	<b>WTD prob</b>	18.8	0.039	0.1	2.13	14.4	0.037	0.9	1.97
	<b>90 days</b>	-8.0	0.029	13.7	1.16	-10.2	0.030	3.5	1.26
	<b>30 days</b>	-13.4	0.038	3.4	2.09	-15.2	0.042	2.6	2.50
	<i>1:10 CC</i>								
	<b>True expo</b>	0.0	0.036	94.6	1.00	0.1	0.036	95.3	1.00
	<b>CC WTD</b>	1.4	0.038	92.9	1.11	1.7	0.039	92.7	1.14
	<b>WTD prob</b>	8.5	0.048	49.2	1.71	5.9	0.046	71.5	1.63
	<b>90 days</b>	-9.9	0.038	18.4	1.10	-12.8	0.039	4.7	1.17
	<b>30 days</b>	-20.0	0.048	0.4	1.74	-22.1	0.052	0.2	2.05

2

3 The datasets had a sample size of 39,600, and on average 80% of patients continue treatment at the  
 4 index date, 25% of patients have a prescription redemption in the year before the index date and the  
 5 true OR is 3. For each setting 2,500 datasets were generated and analyzed. The Weibull Backward

1 Recurrence Density used to generate data corresponded to a Weibull distribution with the same  
2 mean and variance as a Log-Normal distribution with  $\mu$  and  $\sigma$  as its parameters, see text for details.  
3

4 (\*) *Analysis methods:*

5 *1:1 CC* indicates 1 control per case, *1:10 CC* indicates 10 controls per case.

6 **True expo** – logistic regression with the actual exposure status as covariate (the reference analysis).  
7 **CC WTD** – estimation based on joint likelihood for case-control status and the reverse WTD, Log-  
8 Normal Backward Recurrence Density.

9 **WTD prob** – a reverse WTD with Log-Normal Backward Recurrence Density is estimated to  
10 predict the probability of an individual being exposed and this exposure probability is used as  
11 covariate in logistic regression.

12 **90 days** – individuals are considered exposed if they have a redemption < 90 days before index  
13 date.

14 **30 days** – individuals are considered exposed if they have a redemption < 30 days before index  
15 date, logistic regression.

1 Table 3. Estimated association between NSAID use and severe upper gastrointestinal bleeding

<b>Exposure definition</b>	<b>n</b>	<b>Crude OR (95% confidence interval)</b>	<b>Adjusted OR (95% confidence interval)</b>	<b>Upper/lower confidence limit ratio for adjusted OR</b>
Fixed window, 30 days	9,453			
<i>Conditional</i>		7.06 (6.17 - 8.06)	5.17 (2.40 - 11.11)	4.62
<i>Ordinary</i>		6.91 (6.19 - 7.71)	3.85 (2.09 - 7.07)	3.38
Fixed window, 90 days	12,662			
<i>Conditional</i>		4.96 (4.46 - 5.51)	4.73 (2.72 - 8.23)	3.02
<i>Ordinary</i>		4.87 (4.43 - 5.36)	3.52 (2.19 - 5.65)	2.58
WTD treatment probability	39,119			
<i>Conditional</i>		6.99 (6.35 - 7.69)	4.37 (3.62 - 5.28)	1.46
<i>Ordinary</i>		6.90 (6.28 - 7.58)	3.94 (3.29 - 4.72)	1.43
Joint likelihood model	39,119	5.57 (5.08 - 6.05)	2.52 (1.59 - 3.45)	2.18

2

3 Case-control study of 3568 cases and 35552 controls. See text for technical description of exposure  
 4 definitions and covariate adjustment. *Conditional* results are based on conditional logistic  
 5 regression which accounts for matching on age and sex. *Ordinary* results are based on ordinary  
 6 logistic regression (unconditional), but where sex and age are included as covariates, both in the  
 7 crude and adjusted analyses.

# Figures

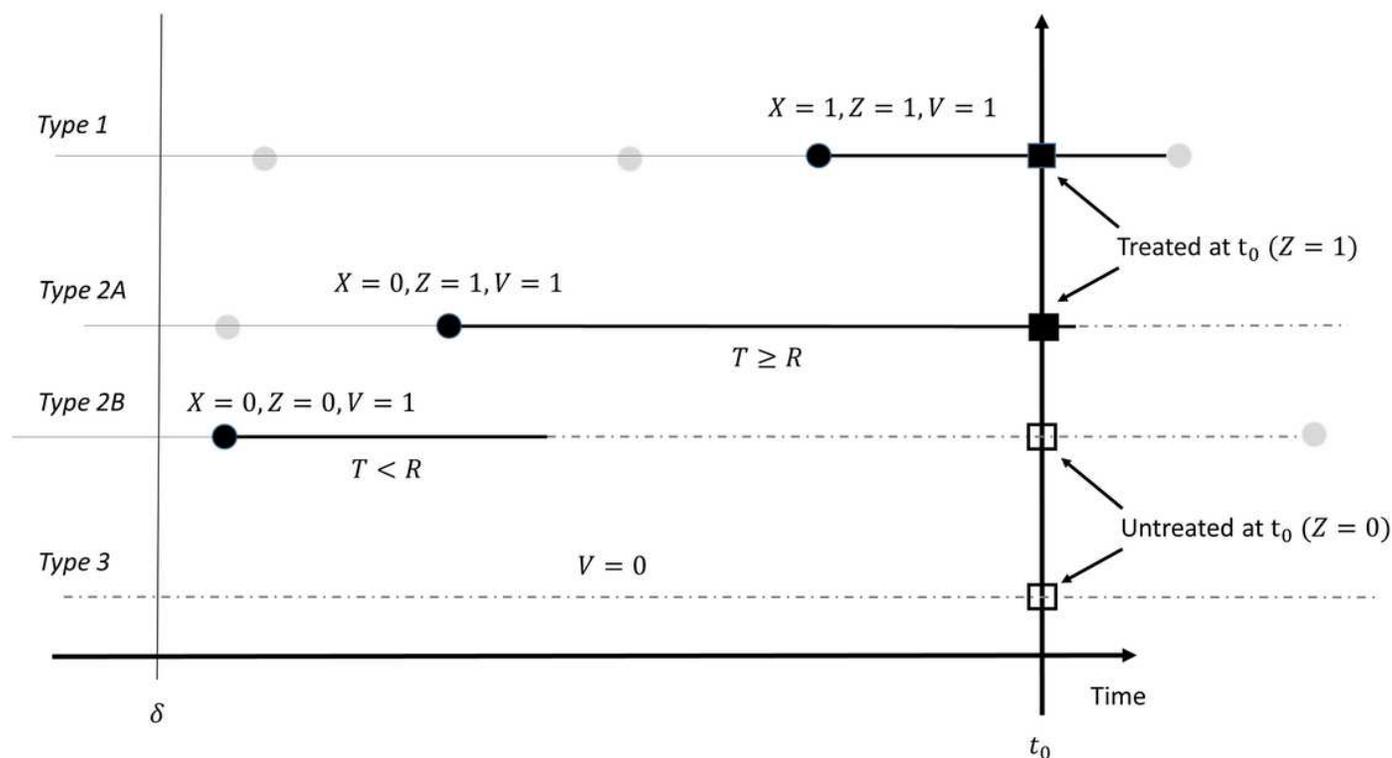


Figure 1

Four typical persons (horizontal lines) representing the different types of contributions to the likelihood with respect to exposure status on the index date indicated by squares (filled: treated, unfilled: untreated). Black bullets represent prescription redemptions included in the likelihood, grey bullets are redemptions not included in the likelihood. The thicker black lines represent periods of treatment, whereas dashed lines are periods without treatment.  $T$  is duration of the prescription and  $R$  is time from the last prescription redemption before the index date  $t_0$  until the index date.  $X$  indicates whether a patient continues treatment,  $Z$  exposure status on the index date, and  $V$  whether a prescription is redeemed in the interval  $(\delta; t_0)$ . For simplicity, index dates have been aligned on the time scale. Type refers to patient type, see text for details.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)