Received: 6 June 2017 | Revised: 21 August 2017 | Accepted: 3 September 2017

DOI: 10.1002/pds.4326

WILEY

ORIGINAL REPORT

Using probability of drug use as independent variable in a register-based pharmacoepidemiological cause-effect study—An application of the reverse waiting time distribution

Jesper Hallas[1,2] | Anton Pottegård[1,3] | Henrik Støvring[4]

[1] Clinical Pharmacology and Pharmacy, University of Southern Denmark, Odense, Denmark

[2] Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark

[3] Hospital Pharmacy, Odense University Hospital, Odense, Denmark

[4] Biostatistics, Department of Public Health, Aarhus University, Aarhus, Denmark

Correspondence
J. Hallas, Clinical Pharmacology and Pharmacy, University of Southern Denmark, JB Winsløwsvej 19, 2, 5000 Odense C, Denmark.
Email: jhallas@health.sdu.dk

Funding information
Aarhus University; University of Southern Denmark

Abstract

Background: In register-based pharmacoepidemiological studies, each day of follow-up is usually categorized either as exposed or unexposed. However, there is an underlying continuous probability of exposure, and by insisting on a dichotomy, researchers unwillingly force a nondifferential misclassification into their analyses. We have recently developed a model whereby probability of exposure can be modeled, and we tested this on an empirical case of nonsteroidal anti-inflammatory drug (NSAID)-induced upper gastrointestinal bleeding (UGIB).

Methods: We used a case-controls data set, consisting of 3568 cases of severe UGIB and 35 552 matched controls. Exposure to NSAID was based on 3 different conventional dichotomous measures. In addition, we tested 3 probabilistic exposure measures, a simple univariate backward-recurrence model, a "full" multivariable model, and a "reduced" multivariable model. Odds ratios (ORs) and 95% confidence intervals for the association between NSAID use and UGIB were calculated by conditional logistic regression, while adjusting for preselected confounders.

Results: Compared to the conventional dichotomous exposure measures, the probabilistic exposure measures generated adjusted ORs in the upper range (4.37-4.75) while at the same time having the most narrow confidence intervals (ratio between upper and lower confidence limit, 1.46-1.50). Some ORs generated by conventional measures were higher than the probabilistic ORs, but only when the assumed period of intake was unrealistically short.

Conclusion: The pattern of high ORs and narrow confidence intervals in probabilistic exposure measures is compatible with less nondifferential misclassification of exposure than in a dichotomous exposure model. Probabilistic exposure measures appear to be an attractive alternative to conventional exposure measures.

KEYWORDS

databases, exposure, misclassification, NSAID, pharmacoepidemiology, upper gastrointestinal bleeding

1 | INTRODUCTION

In register-based studies of drug effects, all days of follow-up are typically classified as either exposed or unexposed, depending on some rule or algorithm that transforms the prescription data into a temporal exposure pattern. The key to this classification is usually the prescription's recency; days long after the latest prescription are classified as unexposed, while days soon after this prescription are classified as exposed. By insisting on classifying all days as either exposed or unexposed, researchers ignore the fact that the prescription data reflect an underlying continuous probability of exposure, not a dichotomy. Inadvertently, researchers thereby force an exposure misclassification into their analysis, classifying a low probability of treatment as nonexposure and a high probability of treatment as exposure. These misclassifications are likely to be nondifferential, thus leading to an attenuation of the association.

Based on theory for renewal processes, we have recently proposed a method whereby the probability of (still) being treated can be modeled as a function of the distance from the latest prescription.[1] We have extended the method to also include multivariable modeling such that the probability function may depend on covariates, such as age, sex, package size, coprescribed medication, and comorbidity.[2]

We propose that the problem of misclassification from using a dichotomous exposure measure can be attenuated by using such a continuous exposure probability as main exposure variable in the regressions. An intuitive rationale is that the odds ratio (OR) associated with a continuous variable is interpreted as the OR associated with an increment of one unit. In this context, it corresponds to the OR comparing a definitely exposed person (probability of exposure, 1) with a definitely unexposed person (probability of exposure, 0).

The aim of this paper is to perform regression on treatment probability and compare its output to results from using conventional approaches to assigning dichotomous exposure status from registry data. As our motivating example, we have chosen the association between traditional nonsteroidal anti-inflammatory drugs (NSAIDs) and upper gastrointestinal bleeding (UGIB). Treatment with NSAID is characterized by a mixed episodic and continuous treatment pattern and by having a well-established and fairly strong association with UGIB.

KEY POINTS

- In register-based pharmacoepidemiological studies, the probability of being treated at a certain point in time is dependent on the time passed since the last prescription for the drug. This probability can be modelled.
- Researchers introduce misclassification by classifying all follow-up as either exposed or unexposed, thereby ignoring an underlying continuous treatment probability.
- We demonstrate how the modelled probability can be used as exposure in multivariable regression.
- Our results suggest that avoiding misclassification with use of continuous treatment probability is statistically more efficient than conventional methods.

2 | METHODS

We analyzed a case-control data set of patients with severe upper gastrointestinal bleeding, using 3 different types of conventional binary exposure methods and 3 different continuous exposure probability models. As an aid in interpretation of the results, we analyzed the time dependency of the OR, relative to the timing of the latest NSAID prescription before the index date.

2.1 | Cases and controls

We used a case material described in detail in previous publications.[3,4] In brief, our source population was the residents of Funen County during 1995 to 2006. We included as potential cases all subjects admitted with a diagnosis compatible with UGIB (n = 12 607). These underwent a manual review of all discharge summaries. During the review, the study group was blinded to the exposure status of potential cases. Each case was assigned an index date defined as the first registered date of an UGIB diagnosis.

Controls were selected by a risk-set sampling strategy, ie, for each case, we randomly selected 10 controls among the subjects in our source population who matched the case by sex and birth year. Controls were assigned an index date identical to the outcome defining date of the corresponding case. We allowed that cases could be selected as controls before they had their case-defining event. Thereby, the calculated OR is an unbiased estimate of the incidence rate ratio that would have emerged from a cohort study, based on the same source population.

We required that both cases and controls had been residents of Funen for at least 1 year on the index date. As some of the very old cases had less than 10 eligible controls, the final control to case ratio deviated slightly from 10:1.

2.2 | Exposure

For all conventional, dichotomous exposure definitions, exposure started on the day of dispensing and lasted until the end of the prescription duration, unless another prescription occurred before, in which case the exposure clock was reset. For continuous exposure probability, we assumed that exposure started on the day of dispensing and restarted the exposure probability function with each new prescription, regardless of when it occurred.

The following exposure algorithms were used:

Dichotomous exposure

- Fixed window: All prescriptions were assigned durations of 30 days. Analyses with 60, 90, and 120 days were used as well.
- Fixed daily amount: All prescriptions duration were calculated under the assumption of a daily intake of 0.2 defined daily doses (DDD).[5] Analyses using 0.5, 1.0, and 1.5 DDD/day were used as well. The number of days assigned to each prescription was thus the amount dispensed with that prescription, measured in DDD, and divided by 0.2, 0.5, 1.0, and 1.5, respectively.
- Simple waiting time distribution (WTD) model: The WTD was used to model the prescription duration, using the method described by Pottegård and Hallas.[6] The WTD percentile was set to 0.75 with additional analyses using 0.80, 0.85, and 0.90.

Continuous exposure probability

- A single, one-size-fits-all probability function for all NSAID dispensings, without consideration of covariates: In this model, we estimated a single reverse WTD for all the controls with the index date as the right-hand end point of the observation window. We used the prior year (365 days) as observation window (thereby ignoring all controls without an NSAID dispensing during the year before their index date). Note that the observation window does not occur as a single period in calendar time, but rather is located relative to the index date. Since the index date for controls can be considered random with respect to dispensing dates, their time

from last dispensing to the index date will follow a reverse WTD. Based on obtained estimates and the date of their last NSAID dispensing, we calculated the probability of being exposed on their index date for all controls and cases. If a subject did not have any observed NSAID dispensing, we set their probability of being exposed to zero.

The rationale for choosing controls for modeling continuous exposure probability was that for cases, the outcomes could potentially interfere with the pattern of prescription renewal, for example, by truncating treatment episodes.

- A tailored multivariable probability function. We used the same approach as above, except that we estimated parameters of the reverse WTD, which depended on the following covariates: age, sex, quantity dispensed (in DDD), use of ibuprofene (the dominant NSAID), concurrent use of proton pump inhibitors, a diagnosis of rheumatoid arthritis, psoriasis arthritis or spondylarthritis, and concurrent use of methotrexate or systemic corticosteroids. The latter variables (use of proton pump inhibitors and onwards) were intended to reflect markers of long-term NSAID use. Based on estimated parameters, observed covariates and date of last NSAID dispensing before the index date, we calculated the probability of being exposed on their index date for each case and control. Apart from choice of ibuprofene, which accounts for about 60% of the total sales, we did not incorporate the chosen NSAID substance in the model.
- In a final, reduced model, we only included the covariates that reached statistical significance in the previous model: sex, age, dispensed quantity, and concurrent use of methotrexate or systemic corticosteroids. This model was included to examine how much of the predictive ability was lost by exclusion of apparently unimportant predictors (ie, by omission of use of ibuprofene, concurrent use of proton pump inhibitors, and a diagnosis of rheumatoid arthritis, psoriasis arthritis, or spondylarthritis).

In all three continuous models, the probability of being exposed was set to zero, when no NSAID dispensings had ever been observed for a subject. A dedicated Stata package for estimating the exposure probability by any of these models (wtdttt) can be downloaded from the IDEAS repository (http://ideas.repec.org) and may be installed in Stata using a search for the package name, ie, –search wtdttt, all–.

2.3 | Data analysis

By using conditional logistic regression, we estimated the crude and adjusted ORs with 95% confidence intervals (CIs) for the association between use of NSAIDs and UGIB. Confounding by age, sex, and calendar time was accounted for by the matching and conditional analysis. For the adjusted ORs, the following potential confounders were included: (1) current use of the following drugs: vitamin K antagonists, aspirin, other antiplatelet drugs, dipyridamol, beta-blockers, selective serotonin reuptake inhibitors, systemic corticosteroids, proton pump inhibitors, H2 receptor antagonists, statins, nitrates, spironolactone, calcium antagonists, and bisphosphonates; (2) any history of the following events: previous UGIB, *Helicobacter pylori* eradication, peptic ulcer, chronic obstructive pulmonary disease, diabetes, ischemic heart disease, heart failure, stroke, hypertension, inflammatory bowel disease, malignant disease, and renal failure; and (3) prescription or diagnosis markers of smoking or excessive alcohol consumption. For all drugs used as covariates, current drug use was defined by the redeeming of a prescription within less than 120 days before the index date.

We used the upper/lower confidence limit ratio (ULCLR) of the adjusted OR as a measure of statistical precision in all analyses. This is equivalent to considering the magnitude of the standard error on the log-odds scale, which is a measure of statistical efficiency.

To describe the relationship between dichotomous and continuous exposure measures, we calculated selected percentiles of the exposure probability for the simple WTD model, the full-model multivariable WTD, and the reduced-model multivariable WTD for persons, who were deemed either exposed or unexposed by each of the conventional exposure methods. For subjects deemed exposed according to conventional binary measures, we calculated the 25, 50, and 75 percentiles, and for unexposed, the 90, 95, and 99 percentiles. For subjects deemed unexposed by conventional binary measures, the probabilities of exposure according to the WTD models were generally quite low, and describing their distribution by the 25, 50, and 75 percentiles would be uninformative.

As an aid in interpreting the main results, we also characterized the association between NSAID use and UGIB as a function of the time since last NSAID prescription. In this analysis, we categorized all cases and controls with respect to their time since last prescription in 10-day intervals. For a given time-band, say 70 to 79 days, we included only cases and controls who had either no prescriptions within the latest year (reference) or had their latest prescription exactly within the interval of 70 to 79 days before the index date. All others were disregarded when analyzing this particular time-band. A conventional crude analysis was then performed on this restricted material. As the same reference was used for all analyses, the ORs generated for each time-band are mutually comparable. Some of the strata were quite thin, and we thus refrained from including other covariates than NSAID use in this analysis. The effects of age, sex, and calendar time were handled by the matching.

For information on codes (ICD10 and ATC) used to define the covariates, see Appendix S1. We used Stata v 14 for all analyses. According to Danish law, review by an ethics committee is not required for purely register-base studies.[7]

3 | RESULTS

We identified 3571 manually validated cases and 35 582 controls. Their median age was 75 (interquartile range, 64-83), and 50.7% were male. All included comorbidities and currently used drugs were more common among cases than controls, as was the use of ulcerogenic medications (Table 1). For cases, 2610 (73%) were ever-users of NSAIDs, while corresponding figures for controls were 21 740 and 61%.

The results of the regression analyses are shown in Table 2. For all conventional approaches, the OR was higher with a short exposure period assigned to each prescription than with a long. Short exposure periods were also associated with lower counts of exposed cases and controls, and with wider confidence intervals for the OR (i.e. higher

TABLE 1 Characteristics of cases and controls

	Cases n = 3568	Controls n = 35 552
Demographics		
Age, median (IQR)	75 (64-83)	75 (64-83)
Male sex	1811 (50.7%)	18 029 (50.7%)
Current drug use		
VKA	183 (5.1%)	823 (2.3%)
ASA	696 (19.5%)	3436 (9.7%)
Other antiplatelet drugs	197 (5.5%)	782 (2.2%)
SSRI	429 (12.0%)	2038 (5.7%)
Systemic corticosteroids	384 (10.8%)	1638 (4.6%)
PPI	521 (14.6%)	2037 (5.7%)
H2 receptor antagonists	294 (8.2%)	958 (2.7%)
Statins	237 (6.6%)	1572 (4.4%)
Nitrates	318 (8.9%)	1678 (4.7%)
Spironolactone	208 (5.8%)	599 (1.7%)
Calcium antagonists	588 (16.5%)	3829 (10.8%)
Bisphosphonates	70 (2.0%)	439 (1.2%)
History of		
UGIB	95 (2.7%)	175 (0.5%)
HP eradication	160 (4.5%)	467 (1.3%)
Peptic ulcer	218 (6.1%)	535 (1.5%)
COPD	256 (7.2%)	1044 (2.9%)
Diabetes	404 (11.3%)	2167 (6.1%)
Ischemic heart disease	867 (24.3%)	5272 (14.8%)
Heart failure	279 (7.8%)	1164 (3.3%)
Stroke	353 (9.9%)	1835 (5.2%)
Hypertension	412 (11.5%)	1863 (5.2%)
Inflammatory bowel disease	23 (0.6%)	107 (0.3%)
Malignant disease	244 (6.8%)	1711 (4.8%)
Renal failure	94 (2.6%)	205 (0.6%)
Alcohol-related markers	166 (4.6%)	336 (0.9%)
Tobacco-related markers	1148 (32.1%)	8364 (23.5%)

Abbreviations: ASA, acetylsalicylic acid; COPD, chronic obstructive pulmonary disease; HP, *Helicobacter pylori*; IQR, interquartile range; PPI, proton pump inhibitor; SSRI, selective serotonin reuptake inhibitor; UGIB, upper gastrointestinal bleeding; VKA, vitamin K antagonist.

Cases with severe upper gastrointestinal bleeding in Funen County 1999-2006, with their controls, matched by sex and birthyear.

ULCLR). The narrowest confidence intervals were found for the probabilistic exposure measures (ULCLR range 1.46–1.50 vs 1.62–5.06 for the dichotomous exposure measures). There was little difference between ORs generated by the 3 probabilistic measures (ORs, 4.37-4.75).

The relationship between dichotomous and continuous exposure measures is demonstrated in Table 3. As expected, the conventional measures that assumed a quick intake of the dispensed medication (eg, 30-day fixed window) had high values for the probabilistic exposure measures, indicating a high positive predictive value. However, the values were also fairly high for the unexposed, indicating that a person who was deemed unexposed by assuming a 30-day window had a substantial probability of being exposed according to the probabilistic model, ie, a low sensitivity for the 30-day window. There was very little difference between the probabilities from the full multivariable probabilistic model, compared to the reduced model.

The temporal pattern of the OR, relative to the time since last NSAID prescription is demonstrated in Figure 1. As can be seen, the OR quickly decreases, starting at 7.6 (CI, 6.8-8.6) for day 1 to 9, reaching 1.7 (CI, 1.3-2.3) at the interval 90 to 99 days and 1.4 (CI, 0.9-2.1) at 180 to 189 days.

4 | DISCUSSION

The 3 analyses with probabilistic exposure measures produced ORs almost as high as the highest dichotomous analyses and with more narrow confidence intervals. While it is a relevant benchmark for comparison of exposure definition to consider the strength of the estimated effect of NSAIDs on UGIBs—increasing misclassification dilutes the association—the width of confidence intervals must also be considered. A possible downside of the probabilistic model is its complexity, but this is to some extent mitigated by the observation that complex multivariable models did not perform better than simpler models. In addition, we have published a Stata package to aid its implementation.

Our rationale for using a continuous probability of exposure was to reduce the error caused by nondifferential misclassification. We would thus expect the continuous measures to produce high ORs. Our results did confirm this, although a few of the ORs produced by conventional exposure measures were higher. This should, however, not be viewed as a failure of our rationale. As shown in Figure 1, there is a strong dependency of the OR on the time since last NSAID prescription, above and beyond what can be explained by a decreasing likelihood of being exposed. This time dependency in OR may be explained by depletion of susceptibles[8] or possibly by a genuine biologic adaptation towards NSAID exposure.[9] With depletion of susceptibles, subjects who have had the outcome of interest are continuously removed from the population of drug users, whereby a selected group that is tolerant to the adverse drug effect remains.[10] Thereby, exposure measures that only categorize the subjects as exposed the first few days after a new prescription, eg, by using a short fixed time window or assuming a high daily intake, are likely to produce high ORs. However, these extreme exposure measures, which by all accounts are unrealistic in our setting, all have fewer exposed subjects and thereby considerably less precision. If the objective is to establish an association, our continuous measures seem more attractive, as they tended to produce fairly high ORs, but with much more narrow confidence intervals than the dichotomous exposure measures.

One limitation of our study is the time dependency of the association, which renders our interpretation difficult. However, examples with immediate, transient effect, strong associations, and low level of time dependency are difficult to find. The time dependency has a strong component of depletion of susceptibles, which again is explained by the mere presence of variability in sensitivity of treated subjects towards the adverse outcome. It is difficult to think of examples of adverse drug reactions without between-subject variability in sensitivity. Thus, time dependency of OR for a transient effect while being treated is almost universal. Another limitation is that we did not have data on actual daily dose of NSAID, which is likely both a

TABLE 2 Crude and adjusted OR for an association between NSAID use and severe upper gastrointestinal bleeding, according to different exposure definitions

Exposure Definition	Exposure Probability, Cases	Exposure Probability, Controls	Crude OR (95% Confidence Interval)	Adjusted OR[b] (95% Confidence Interval)	Upper/Lower Confidence Limit Ratio for Adjusted OR
Dichotomous exposure					
Fixed window					
30 d	45.0%	10.8%	7.06 (6.17-8.06)	5.17 (2.40-11.11)	4.62
60 d	52.4%	16.3%	5.78 (5.16-6.47)	5.13 (2.75-9.55)	3.47
90 d	55.3%	20.3%	4.96 (4.46-5.51)	4.73 (2.72-8.23)	3.02
120 d	56.5%	22.9%	4.44 (4.01-4.91)	3.64 (2.14-6.18)	2.89
Fixed daily intake					
1.5 DDD/d	41.5%	9.1%	7.42 (6.42-8.57)	6.48 (2.88-14.57)	5.06
1.0 DDD/d	47.8%	11.9%	6.90 (6.08-7.83)	5.95 (3.02-11.71)	3.88
0.5 DDD/d	52.9%	16.7%	5.75 (5.14-6.44)	2.78 (1.77-4.37)	2.47
0.2 DDD/d	56.6%	23.2%	4.43 (4.00-4.90)	1.49 (1.16-1.93)	1.67
Simple WTD model					
0.75	56.0%	21.5%	4.75 (4.28-5.26)	4.26 (2.49-7.28)	2.92
0.80	56.5%	22.8%	4.47 (4.03-4.95)	3.72 (2.18-6.33)	2.90
0.85	57.2%	24.3%	4.17 (3.77-4.60)	1.64 (1.19-2.27)	1.90
0.90	58.1%	26.3%	3.90 (3.54-4.30)	1.36 (1.07-1.72)	1.62
Continuous treatment probability					
Simple model	0.057 (<0.001-0.915)[a]	<0.001 (<0.001-0.027)[a]	6.77 (6.16-7.45)	4.75 (3.88-5.83)	1.50
Full multivariable model	0.037 (<0.001-0.903)[a]	<0.001 (<0.001-0.014)[a]	6.99 (6.35-7.69)	4.37 (3.62-5.28)	1.46
Reduced multivariable model	0.038 (<0.001-0.895)[a]	<0.001 (<0.001-0.014)[a]	6.98 (6.34-7.68)	4.46 (3.69-5.39)	1.46

Abbreviations: DDD, defined daily dose; NSAID, nonsteroidal anti-inflammatory drug; OR, odds ratio; WTD, waiting time distribution.

Case-control study of 3568 cases and 35 552 controls. See text for technical description of exposure definitions.

[a]Median and interquartile range for ever-users of NSAIDs.

[b]Adjusted for current use of the following drugs: vitamin K antagonists, aspirin, other antiplatelet drugs, dipyridamol, beta-blockers, selective serotonin reuptake inhibitors, systemic corticosteroids, proton pump inhibitors, H2 receptor antagonists, statins, nitrates, spironolactone, calcium antagonists, bisphosphonates, any history of the following events, non-variceal upper GI bleeding, *Helicobacter pylori* eradication, peptic ulcer, chronic obstructive pulmonary disease, diabetes, ischemic heart disease, heart failure, stroke, hypertension, inflammatory bowel disease, malignant disease, renal failure; and prescription or diagnosis markers of smoking or excessive alcohol consumption.

determinant of the outcome and related to the interval between dispensings, ie, the continuous exposure probability. Unfortunately, the prescribed dose is not recorded in our data source. As a crude proxy, we have incorporated the dispensed quantity in the multivariable dichotomous exposure models, assuming that persons with high doses also have large quantities dispensed.

In our application, we have assigned treatment probabilities based on the last observed redemption before the index date, and by assuming that the subsequent inter-arrival time to the next redemption is a randomly chosen one. For both cases and controls, this is, however, not likely to be the case, but for different reasons. For controls, the index date is evidently chosen independently of their prescription renewal dates. Consequently, the intercepted interval will suffer from length bias, since index dates will have a higher probability of hitting longer inter-arrival times between prescriptions. For cases, index date and date of last prescription are in contrast very likely to be dependent, in particular when the risk of UGIBs varies with time since treatment initiation with NSAIDs. The use of covariates to model the inter-arrival distribution can be expected to reduce the impact of this, but we did not see major differences between models with and without covariates.

Our approach is not the only example where likelihood of exposure is used as input for a regression. In some variants of the instrumental variable technique, the "instrument" may represent a preference for treating patients with a given drug. This preference is a determinant of exposure without any independent effect on the outcome.[11] However, in an instrumental variable analysis, the actual exposure data are deliberately left out of the regression. Unlike our approach, instrumental variable analysis often results in a substantial loss of precision, because the association between instrument and exposure is weak.[12] Furthermore, the purpose of an instrumental variable approach is to deal with confounding, not misclassification as in our approach.

In reality, patients are on any given day treated or not depending on whether they took the drug that day. Unfortunately, this true status is not observable in pharmacoepidemiologic databases. It has therefore been a standard endeavor in pharmacoepidemiology to emulate this status from prescription data, although with mixed results and little consensus on optimal strategies. As our results show, the estimated risk varies considerably with the definition used to classify treatment status. In a sense, the approach we have suggested here is more pragmatic. Using the reverse WTD, we are able to estimate the inter-arrival distribution of time between 2 subsequent redemptions of users in continued treatment. We then interpret one minus this cumulative distribution function as the probability of a patient still being treated. Just after a prescription redemption, this yields a

TABLE 3 Distribution of estimated exposure probability among subjects who are classified as exposed, according to simple dichotomous criteria

	Continuous Treatment Probability, Simple Model		Continuous Treatment Probability, Full Multivariable Model		Continuous Treatment Probability, Reduced Multivariable Model	
	Exposed[a]	Unexposed[b]	Exposed[a]	Unexposed[b]	Exposed[a]	Unexposed[b]
Fixed window						
30 d	0.985 (0.926-0.999)	0.350 (0.062-0.760)	0.976 (0.876-0.999)	0.295 (0.036-0.864)	0.974 (0.870-0.999)	0.296 (0.036-0.861)
60 d	0.915 (0.732-0.995)	0.125 (0.015-0.394)	0.905 (0.640-0.994)	0.073 (0.008-0.560)	0.900 (0.635-0.994)	0.074 (0.008-0.557)
90 d	0.829 (0.509-0.985)	0.043 (0.005-0.203)	0.803 (0.426-0.982)	0.023 (0.003-0.314)	0.806 (0.424-0.981)	0.024 (0.003-0.313)
120 d	0.760 (0.367-0.980)	0.020 (0.003-0.104)	0.717 (0.289-0.971)	0.011 (0.002-0.189)	0.713 (0.292-0.968)	0.011 (0.002-0.186)
Fixed daily intake						
1.5 DDD/d	0.985 (0.904-1.000)	0.475 (0.095-0.953)	0.990 (0.947-1.000)	0.396 (0.054-0.854)	0.989 (0.946-1.000)	0.393 (0.054-0.849)
1.0 DDD/d	0.968 (0.802-0.999)	0.310 (0.045-0.892)	0.967 (0.869-0.999)	0.218 (0.024-0.716)	0.964 (0.865-0.999)	0.216 (0.025-0.700)
0.5 DDD/d	0.904 (0.621-0.995)	0.128 (0.014-0.732)	0.899 (0.642-0.993)	0.068 (0.008-0.485)	0.895 (0.637-0.993)	0.069 (0.008-0.478)
0.2 DDD/d	0.746 (0.303-0.975)	0.029 (0.003-0.394)	0.708 (0.271-0.969)	0.013 (0.002-0.205)	0.698 (0.273-0.967)	0.013 (0.002-0.204)
Simple WTD						
0.75	0.802 (0.433-0.980)	0.031 (0.004-0.150)	0.764 (0.362-0.977)	0.017 (0.002-0.251)	0.762 (0.360-0.976)	0.017 (0.002-0.247)
0.80	0.774 (0.367-0.980)	0.022 (0.003-0.109)	0.723 (0.295-0.971)	0.012 (0.002-0.190)	0.721 (0.297-0.969)	0.012 (0.002-0.187)
0.85	0.718 (0.296-0.975)	0.013 (0.002-0.069)	0.658 (0.214-0.963)	0.008 (0.001-0.136)	0.652 (0.211-0.961)	0.007 (0.001-0.135)
0.90	0.635 (0.212-0.961)	0.007 (0.001-0.049)	0.573 (0.136-0.950)	0.004 (0.001-0.077)	0.566 (0.135-0.947)	0.004 (0.001-0.080)

Comparison of simple continuous treatment probability (CTP) model, full multivariable CTP model, and reduced multivariable CTP model. The reported median for exposed with a 30-day fixed window of 0.985 for the simple CTP means that among all with a redemption within 30 days before the index date, the median probability of being exposed is 98.5% when predicted from the simple WTD model. Among the corresponding unexposed the 95-percentile of 0.350 means that among subjects without a redemption in the last 30 days before the index date, 95% had a predicted probability of being treated smaller than 35%.

[a]Median and interquartile range.

[b]95-, 90- and 99-percentile.

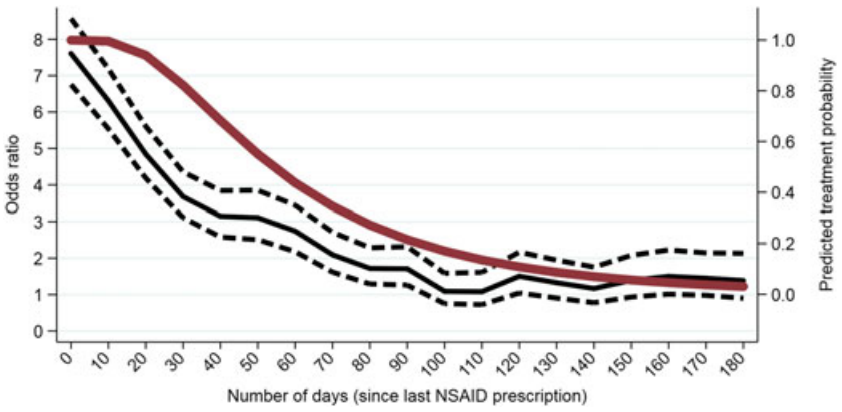

FIGURE 1 Dependency of odds ratio for upper gastrointestinal bleeding as a function of time since last NSAID prescription and the corresponding treatment probability function. Based on 3568 cases and 35 552 controls, sampled from Funen County, Denmark. The dashed lines indicate the 95% confidence *intervals* for the odds ratio. Curves are aligned visually so that they have the same height and that a treatment probability of 0 corresponds to an odds ratio of 1

probability of one, whereas it approaches zero, when sufficient time has passed that all patients in continued treatment would have renewed their prescription. For treatments such as NSAIDs, this assignment of probability may not be optimal in the sense that some patients may use all pills of a redeemed prescription and then stop, whereas others with the same time between redemptions may have been treated periodically in between them. However, we do not think the data allows us to make such distinctions.

We have argued that our approach is attractive as it theoretically bypasses the need for binary classification of exposure. This can be expected to reduce misclassification and thereby reduce an element of conservative bias. We have applied it to an empirical case, essentially producing what we had expected. However, some uncertainties remain that could be the focus of future research. First, our model's coverage is unknown and, given the time dependency of the outcome, difficult to conceptualize. It would thus seem obvious to test our approach in a simulation study, where, for example, we could specify no time dependency. Second, it is conceivable that both exposure and outcome could be modeled in the same procedure, thereby eliminating some of the variability emerging from first developing an exposure model and then carrying the result of this into an outcome model. Again, a simulation study with its full control over specifications would seem attractive. Finally, we would need more empirical experience with our probabilistic exposure model to eventually be able to establish its place in the pharmacoepidemiological armamentarium.

ACKNOWLEDGEMENTS

This study has been funded by the University of Southern Denmark and Aarhus University.

CONFLICT OF INTEREST

None declared.

ORCID

Jesper Hallas http://orcid.org/0000-0002-8097-8708
Anton Pottegård http://orcid.org/0000-0001-9314-5679
Henrik Støvring http://orcid.org/0000-0002-5821-2351

REFERENCES

1. Støvring H, Pottegård A, Hallas J. Determining prescription durations based on the parametric waiting time distribution. *Pharmacoepidemiol Drug Saf*. 2016;25(12):1451-1459.
2. Støvring H, Pottegård A, Hallas J. Refining estimates of prescription durations by using observed covariates in pharmacoepidemiological databases: an application of the reverse waiting time distribution. *Pharmacoepidemiol Drug Saf*. 2017. https://doi.org/10.1002/pds.4216
3. Dall M, Schaffalitzky de Muckadell OB, Lassen AT, Hansen JM, Hallas J. An association between selective serotonin reuptake inhibitor use and serious upper gastrointestinal bleeding. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc*. 2009;7(12):1314-1321.
4. Gulmez SE, Lassen AT, Aalykke C, et al. Do statins protect against upper gastrointestinal bleeding? *Br J Clin Pharmacol*. 2009;67(4):460-465.
5. www.whocc.no
6. Pottegård A, Hallas J. Assigning exposure duration to single prescriptions by use of the waiting time distribution. *Pharmacoepidemiol Drug Saf*. 2013;22(8):803-809.
7. Thygesen LC, Daasnes C, Thaulow I, Brønnum-Hansen H. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scand J Public Health*. 2011;39(7 Suppl):12-16.
8. Moride Y, Abenhaim L, Yola M, Lucein A. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol*. 1994;47(7):731-737.
9. Alderman BM, Giraud AS, Yeomans ND. Adaptation of the stomach to injury from nonsteroidal anti-inflammatory drugs. *Curr Gastroenterol Rep*. 2001;3(1):1-2.
10. Renoux C, Dell'Aniello S, Brenner B, Suissa S. Bias from depletion of susceptibles: the example of hormone replacement therapy and the risk of venous thromboembolism. *Pharmacoepidemiol Drug Saf*. 2017;26(5):554-560.
11. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Instrumental variable analysis of secondary pharmacoepidemiologic data. *Epidemiol Camb Mass*. 2006;17(4):373-374.
12. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiol Camb Mass*. 2006;17(3):268-275.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Hallas J, Pottegård A, Støvring H. Using probability of drug use as independent variable in a register-based pharmacoepidemiological cause-effect study—An application of the reverse waiting time distribution. *Pharmacoepidemiol Drug Saf*. 2017;1–7. https://doi.org/10.1002/pds.4326