PHARMACOEPIDEMIOLOGY AND DRUG SAFETY (2017)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.4216

ORIGINAL REPORT

Refining estimates of prescription durations by using observed covariates in pharmacoepidemiological databases: an application of the reverse waiting time distribution

Henrik Støvring[1]*, Anton Pottegård[2] and Jesper Hallas[2]

[1] *Biostatistics, Department of Public Health, Aarhus University, Aarhus, Denmark*
[2] *Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, Odense, Denmark*

ABSTRACT

Purpose The study aimed to develop an automated method to estimate prescription durations in pharmacoepidemiological studies that may depend on patient and redemption characteristics.

Methods We developed an estimation algorithm based on maximum likelihood estimation for the reverse waiting time distribution (WTD), which is the distribution of time from the last prescription of each patient within a time window to the end of the time window. The reverse WTD consists of two distinctly different components: one component for prevalent users and one for patients stopping treatment. We extended the model to allow parameters of the reverse WTD to depend on linear combinations of covariates to obtain estimates and confidence intervals for percentiles of the inter-arrival density (time from one prescription to the subsequent). We applied the method to redemptions of warfarin, using the amount of drug filled, patient sex and patient age as covariates.

Results The estimated prescription durations increased with redeemed amount and age. Women generally had longer prescription durations, which increased more with age than men. For 70-year-old women redeeming 300+ pills, we predicted a 95th percentile of the inter-arrival density of 225 (95%CI: 201, 249) days. For 50-year-old men redeeming 100 pills, the corresponding prediction was 97 (88, 106) days.

Conclusions The algorithm allows estimation of prescription durations based on the reverse WTD, which can depend upon observed covariates. Statistical uncertainty intervals and tests allow statistical inference on the influence of observed patient and prescription characteristics. The method may replace *ad hoc* decision rules. Copyright © 2017 John Wiley & Sons, Ltd.

KEY WORDS—prescription durations; covariates; reverse waiting time distribution; maximum likelihood; parametric modelling; pharmacoepidemiology

Received 17 October 2016; Revised 28 February 2017; Accepted 31 March 2017

INTRODUCTION

Determining prescription durations is fundamental to several types of studies in pharmacoepidemiology. In drug utilization studies, patients are classified as being incident, prevalent or having stopped with treatment depending on their prescription redemptions and the assumed durations of individual redemptions.[1–3] In studies on effects and side effects of drugs, exposure to the drug must similarly be determined based on prescription redemptions and assumed durations of individual redemptions.[4]

Usually, criteria for determining prescription durations in pharmacoepidemiology are *ad hoc* decision rules based on expert insights with some input from observed treatment patterns. Although the topic has received substantial interest in pharmacoepidemiological research, there is a scarcity of validated methods and little agreement on the optimal approach, despite several studies comparing them.[5–7] In 2013, Pottegård and Hallas suggested to use the waiting time distribution (WTD) to determine durations in an automated manner by estimating a percentile of when a specified percentage of prevalent users had renewed their prescription. They did, however, not consider the distinction between the component distribution observed in the WTD, the so-called forward recurrence distribution, and the

*Correspondence to: H. Støvring, Biostatistics, Department of Public Health, Aarhus University, Bartholins Allé 2, Aarhus C, 8000 Aarhus, Denmark. E-mail: stovring@ph.au.dk

inter-arrival distribution (IAD).[8] It is the latter, which is of primary interest as it describes the distribution of time from one prescription redemption of a patient to the next redemption. Making the distinction between the forward recurrence distribution and the IAD, Støvring *et al*. (2016) showed that the parametric WTD could be used for valid maximum likelihood estimation of a specified percentile of the IAD, and with some robustness against misspecification.[9] Because the method was based on the ordinary WTD, it could not account for individual covariates thought to be predictive for the length of the interval between one prescription and the one following, such as amount of tablets filled or patient characteristics. There is thus a lack of automated methods, which are able to account for patient and redemption characteristics.

In a companion paper to the present paper, we show how the reverse WTD can be analyzed analogously to the ordinary WTD.[10] The amount of tablets obtained when filling a prescription analyzed in the ordinary WTD does not inform us about the waiting time from the start of the time window until that prescription – instead, it is governed by the size of the last prescription before the time window. By contrast, information on covariates of interest will in the reverse WTD pertain directly to the intervals whose distribution we want to estimate (the waiting time from the last prescription redemption to the end of the time window), and this is the fundamental fact exploited in the present paper. With parameter estimates for the effects of covariates on the IAD, we can estimate different percentiles of the IAD, which are relevant for women and men, respectively; for different ages; and for different amounts of the drug redeemed, say. Further, as maximum likelihood estimation is a valid statistical estimation procedure with known asymptotic properties, it provides confidence intervals and statistical tests using standard formulas. The estimated percentiles can then be used in subsequent analyses to define prescription durations such that if a patient has not redeemed a new prescription within the specified percentile, the user is considered to have stopped treatment.

The objective of the present paper is to show how covariates may be included in the parametric model for the reverse WTD, as this will allow estimation of IADs, which depend on covariates. First, we review basic formulas based on renewal process theory, which are used to establish maximum likelihood estimation of the reverse WTD. We then show how regression equations can be added to the model such that the effect of covariates on parameters can be estimated. This allows us to estimate prevalence fractions and IADs, which depend on covariates. We then show how percentiles of the IAD can be estimated conditional on covariate values, as these are of interest when considering prescription durations. We finally show how the method can be applied to estimate prescription durations of warfarin depending on patients' sex, age and the number of pills they redeem.

METHODS

The definition of the reverse WTD and how it can be analyzed using a simple parametric two-component model are presented in detail in the accompanying paper.[10] In brief, we let R denote the time from each patient's last observed prescription redemption in the interval $(t_{-1}; t_0)$ until the end of the interval, t_0. When the underlying redemption times for continued use form a renewal process and the rate of stopping over the observation window $(t_{-1}; t_0)$, the likelihood contribution for a single observed reverse waiting time, r, can be written as

$$L(r;\eta,\theta) = \eta \cdot g(r;\theta) + \frac{1-\eta}{\delta}$$

where η is the fraction of prevalent users among the observed users in the observation window; δ is the width of the observation window, $t_0 - t_{-1}$; and $g(t;\theta)$ is the backward recurrence density (BRD) for prevalent users, which depends on parameters θ. The BRD is related to the IAD (time from one redemption to the subsequent redemption for prevalent patients) by the formula

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

where F is the distribution function and M is the mean of the inter-arrival times. In the accompanying paper, we found that the Log-Normal distribution in a range of applications provided a better fit than the Weibull, and we will therefore only consider the Log-Normal in detail in this paper. Similar results are, however, available using a Weibull distribution. The Log-Normal BRD has the following form:

$$g(r) = \frac{1}{M}\Phi\left(\frac{\log r - \mu}{\sigma}\right)$$

where Φ is the cumulative standard normal

DOI: 10.1002/pds

distribution function and M is the inter-arrival mean given by

$$M = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

Substituting the expression of g into the likelihood, we see that the likelihood depends on three parameters (η, μ, σ). To improve convergence and stability of the maximum likelihood estimation procedure, we logit-transformed the parameter η and log-transformed the parameter σ following previous implementations.[11,12]

We now allow the parameters η, μ, σ—after suitable transformation—to depend on covariates with standard regression notation:

$$\text{logit}\, \eta = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots$$

$$\mu = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \cdots$$

$$\log \sigma = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \cdots$$

The first equation may be thought of as applying logistic regression to model the log-odds of being a prevalent user (logit η). Thus, exp α_0 is the odds of a reference person being prevalent (one for whom all x's are zero) and exp α_1 is the odds ratio associated with a one unit change in x_1. The next two equations correspond to what we would have performed using linear regression on a sample of log-transformed inter-arrival times from the prevalent users. A convenient interpretation of exp β_0 is therefore that it represents median time to the next prescription for a reference person (a person for whom all y's are zero) and exp β_1 is the factor one should multiply the median with when y_1 is increased by one unit. A more detailed explanation of the interpretation of the parameters is provided in the Appendix, Section A1.

While regression parameters may be of interest in their own right, estimates of inter-arrival percentiles will often be the target parameters. These may be estimated straightforwardly using the standard delta rule. In Stata, this is implemented in the -`predict`- and -`nlcom`- commands.

APPLICATION

Data were obtained for the Region of Southern Denmark (1.2 million inhabitants) where prescription redemptions are captured in Odense PharmacoEpidemiological Database.[13] We extracted all redemptions of warfarin (ATC: B01AA03) for 2014. As we used the reverse WTD, we only considered time from each patient's last redemption before 1 January 2015 to the end of the year 2014. The indication, dosage and refill instruction are not recorded in Odense PharmacoEpidemiological Database, and thus, the time of when patients actually stop treatment is not directly observable. This is even more so, because in Denmark, no upper limit exists for the amount of a drug, which can be prescribed in a single prescription. To overcome this and provide credible estimates of prescription durations, we used the following covariates: number of pills dispensed at a single occasion (categorized as 100, 200 and 300+), patients' age (categorized as 0–49, 50–64, 65–79 and 80+ years); and patients' sex. We first conducted univariate analyses on each of the covariates, before we included all covariates in all three parameter equations in a joint model. In a third and final model, we included an interaction term between age (continuous) and sex together with age (continuous), sex and number of pills redeemed. The continuous age covariate was centred at 50 years and with decades as unit (the value 1 corresponds to 60 years, 2 to 70 years). Because models can become complex and difficult to interpret, we finally show how the model can be used to provide predictions of percentiles of the IAD of interest for different combinations of covariate values.

All statistical analyses were conducted in Stata 14.1.[14] A dedicated software package (`wtdttt`) implementing the method is provided at the IDEAS repository (http://ideas.repec.org) and may be installed in Stata using a search for the package name, that is, –`search wtdttt, all`–.

RESULTS

The characteristics of the 21 090 patients included in the analysis are shown in Table 1.

Figure 1 shows the observed reverse WTDs and the fitted densities stratified on the number of pills obtained at the patient's last prescription redemption in 2014. The fitted curves follow the observed distribution rather closely, demonstrating a good fit by the parametric reverse WTD with a Log-Normal backward recurrence distribution. A clear pattern towards longer durations with increasing number of pills obtained can be seen, as well as a tendency to a higher fraction of prevalence with increasing number of pills redeemed. Similar plots stratified on sex and age categories, respectively, showed similarly good fits of the parametric reverse WTD.

DOI: 10.1002/pds

Table 1. Characteristics of the population of patients redeeming at least one prescription of warfarin in 2014 as recorded by OPED

	Men	Women	All
n	12 447 (59.0%)	8643 (41.0%)	21 090
Age (years) (median—10th and 90th percentile)	73 (56, 85)	77 (55, 88)	74 (55, 87)
Number of pills redeemed			
100	9313 (56.4%)	7189 (43.6%)	16 502
200	2456 (67.4%)	1187 (32.6%)	3643
300+	678 (71.7%)	267 (28.3%)	945

OPED, Odense PharmacoEpidemiological Database.
The number of pills redeemed refers to the last prescription redemption of each patient in 2014.

These trends were also seen in univariate analyses (Table 2) and were largely maintained in the adjusted analysis, where all three covariates were included in the model.

In the adjusted model, the median of the IAD increased with the number of pills redeemed (median ratio for 200 vs 100 pills: 1.54 (1.46, 1.63); 300 vs 100 pills: 2.19 (1.99, 2.40)). This can be interpreted as patients redeeming 200 pills having a median time to the next prescription that is 1.54 times longer than the median of patients redeeming 100 pills, when they have the same sex and belong to the same age group. As the factor is lower than 2, we can infer that patients redeeming larger packages have a higher average daily intake. It may be a simple matter of convenience for these patients; large packages are dispensed to avoid frequent pharmacy visits. The effect of sex on the median also increased slightly over the univariate model (women versus men: 1.07 (1.02, 1.12)) as it did for age. With respect to the fraction of prevalent

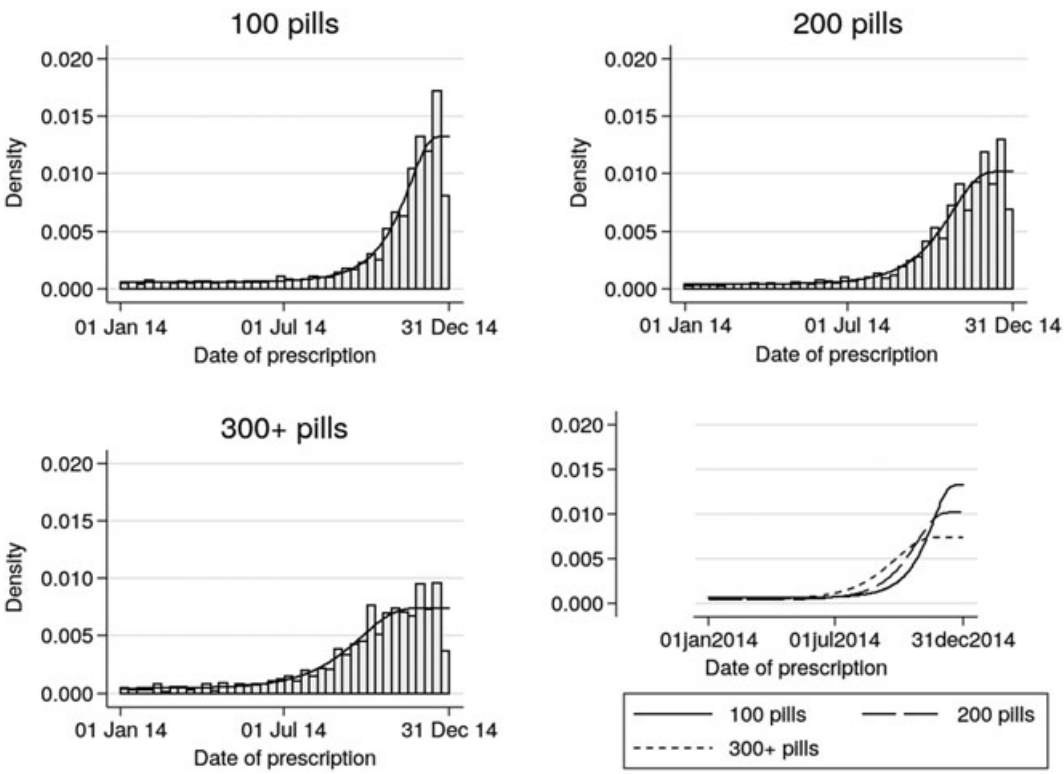

Figure 1. Reverse waiting time distributions for warfarin stratified on number of pills redeemed and with fitted densities using a Log-Normal backward recurrence distribution. The lower right plot compares the three fitted densities shown in the separate histograms. A Log-Normal backward recurrence distribution was used when fitting the WTD. Data from Odense PharmacoEpidemiological Database, 2014

DOI: 10.1002/pds

Table 2. Analyses using the reverse WTD with a Log-Normal backward recurrence distribution with covariates

Model type	Covariate	Covariate value	OR for prevalence ($e^{\alpha_1}, e^{\alpha_2}, \ldots$)	Odds for prevalence (e^{α_0})	Factor for median duration ($e^{\beta_1}, e^{\beta_2}, \ldots$)	Median duration in days (e^{β_0})	Factor for SD on log-time scale ($e^{\gamma_1}, e^{\gamma_2}, \ldots$)	SD on log-time scale (e^{γ_0})
Unadjusted	–	–	–	4.05 (3.82, 4.30)	–	58.8 (57.4, 60.2)	–	0.53 (0.51, 0.55)
Univariate	Sex	Male	1 (Ref)	4.49 (4.14, 4.86)	1 (Ref)	57.7 (55.9, 59.6)	1 (Ref)	0.53 (0.51, 0.56)
		Female	0.79 (0.70, 0.89)	–	1.05 (1.00, 1.10)	–	0.97 (0.90, 1.04)	–
	Age (years)	0–49	0.56 (0.44, 0.72)	–	0.87 (0.76, 1.00)	–	1.05 (0.88, 1.25)	–
		50–64	1 (Ref)	3.88 (3.34, 4.50)	1 (Ref)	48.0 (44.9, 51.4)	1 (Ref)	0.58 (0.53, 0.63)
		65–79	1.41 (1.18, 1.69)	–	1.24 (1.15, 1.33)	–	0.90 (0.81, 0.99)	–
		80+	0.81 (0.68, 0.97)	–	1.41 (1.31, 1.53)	–	0.81 (0.73, 0.91)	–
	Number of pills	100	1 (Ref)	3.61 (3.40, 3.84)	1 (Ref)	54.3 (52.9, 55.8)	1 (Ref)	0.51 (0.49, 0.53)
		200	1.67 (1.38, 2.02)	–	1.46 (1.38, 1.55)	–	0.87 (0.77, 0.97)	–
		300+	1.49 (0.98, 2.27)	–	2.07 (1.89, 2.26)	–	0.75 (0.61, 0.92)	–
Joint	Sex	Female	0.88 (0.78, 0.99)	–	1.07 (1.02, 1.12)	–	0.98 (0.90, 1.06)	–
	Age (years)	0–49	0.56 (0.44, 0.70)	–	0.85 (0.75, 0.96)	–	1.02 (0.84, 1.23)	–
		65–79	1.44 (1.22, 1.71)	–	1.25 (1.17, 1.34)	–	0.91 (0.81, 1.01)	–
		80+	0.90 (0.76, 1.07)	–	1.48 (1.38, 1.60)	–	0.82 (0.73, 0.93)	–
	Number of pills	200	1.60 (1.33, 1.93)	–	1.55 (1.47, 1.63)	–	0.82 (0.74, 0.92)	–
		300+	1.65 (1.05, 2.60)	–	2.19 (2.00, 2.40)	–	0.80 (0.66, 0.96)	–
	Reference	Male, 50–64 years, 100 pills	1 (Ref)	3.49 (3.00, 4.06)	1 (Ref)	42.0 (39.3, 44.9)	1 (Ref)	0.56 (0.51, 0.62)
Final	Sex	Female	0.90 (0.80, 1.01)	–	0.90 (0.80, 1.01)	–	1.31 (1.10, 1.57)	–
	Age (per 10 years)	(age-50)/10	1.12 (1.03, 1.22)	–	1.13 (1.10, 1.17)	–	0.96 (0.90, 1.03)	–
	Interaction(sex and age)	Women, (age-50)/10	0.80 (0.71, 0.91)	–	1.07 (1.02, 1.12)	–	0.88 (0.82, 0.95)	–
	Number of pills	200	1.69 (1.39, 2.05)	–	1.55 (1.47, 1.63)	–	0.84 (0.76, 0.94)	–
		300+	1.55 (1.01, 2.37)	–	2.21 (2.02, 2.42)	–	0.77 (0.65, 0.92)	–
	Reference	Male, 50 years, 100 pills	1 (Ref)	3.02 (2.52, 3.62)	1 (Ref)	39.7 (36.6, 43.0)	1 (Ref)	0.54 (0.46, 0.64)

WTD, waiting time distribution; OR, odds ratio; SD, standard deviation.
Data used are warfarin redemptions recorded in the OPED, 2014.
Estimates are stated with 95% confidence intervals in parentheses.

DOI: 10.1002/pds

patients, changes were small for all three covariates and effects were generally attenuated. In the final model, we used age as a continuous covariate and allowed for its interaction with sex (Table 2). This had virtually no impact on the estimates related to number of pills compared with the model without interactions (median ratio for 200 vs 100 pills: 1.55 (1.47, 1.63); 300 vs 100 pills: 2.21 (2.02, 2.42)). In the final model, it was evident that the effect of age was different among women compared with men for all three parameters. In particular, we found that median duration of inter-arrival times increased faster with age among women than among men (interaction term 1.07 (1.02, 1.12)). For every 10-year difference in age among men, the median duration of inter-arrival times was estimated to increase with a factor 1.13 (1.09, 1.17). No important interaction could be identified between sex and number of pills redeemed (results not shown).

As the models involve three sets of regression equations, it is useful to consider how a model predicts percentiles of interest for the IAD and for different combinations of covariate values. This is reinforced when the model includes interaction terms such as our final model. Table 3 shows the predicted 80th and 95th percentile of the IAD for combinations of age (50 or 70 years), sex (male or female) and number of pills redeemed (100, 200 or 300+). All predictions were based on the final model with sex and age (continuous), their interaction and number of pills redeemed. Remarkably, the predicted 95th percentile for women aged 70 years and redeeming 300+ pills was 225 (201, 249) days, which indicates that a substantial period without redemptions is needed to be reasonably sure that these patients are no longer treated. At the other extreme, men aged 50 years and redeeming 100 pills were predicted to have a 95th percentile of 97 (88, 106) days.

DISCUSSION

We have shown how the reverse WTD may be extended to incorporate covariates in a parametric model. All three basic parameters of the model can be specified via a regression equation, and maximum likelihood estimates of regression coefficients can be obtained. We showed how this could be applied to study prescription durations of warfarin, where we found increasing durations with increasing number of pills redeemed and increasing durations with age, and that the age gradient was steeper for women: older women had longer durations than men.

In pharmacoepidemiology, there has been considerable interest in establishing decision rules for determining prescription durations and evaluating their performance. Classic papers include those of Mantel-Teeuwisse *et al.*[1] and Gardarsdottir *et al.*,[3] who examined the influence of decision rules on the estimated prevalence and duration of treatment episodes. Recently, Meid *et al.*[15] investigated how different standard approaches for defining prescription durations (one tablet per day, defined daily doses, individual longitudinal approximation of the dose) are compared with respect to their ability of estimating the risk of bleeding due to nonsteroidal anti-inflammatory drug use. In a study by Tanskanen *et al.*,[16] a new so-called second generation method was developed to estimate prescription durations based on past individual purchase histories. The approach was evaluated against expert assigned durations for a random sample of patients with relatively high correctness. The method was further evaluated against data on individual drug use obtained through patient interviews in a subsequent study.[17] Our method differs from all these approaches in the important aspect of being based on an explicit and consistent mathematical model. That is, if stopping

Table 3. Predicted percentiles of the inter-arrival distribution for warfarin based on a reverse waiting time distribution with a Log-Normal backward recurrence distribution and with the covariates sex, age as continuous, their interaction and number of pills redeemed (100, 200 and 300+)

			100 pills		200 pills		300 pills	
Sex	Age (years)	Estimated percentile	Estimate(days)	95%CI	Estimate (days)	95%CI	Estimate (days)	95%CI
Male	50	80	62.7	(59.5, 65.8)	90.2	(85.3, 95.0)	125.2	(114.7, 135.7)
Male	70	80	78.1	(76.2, 80.0)	112.9	(108.7, 117.1)	157.1	(145.4, 168.8)
Female	50	80	65.4	(60.8, 70.1)	92.2	(85.1, 99.3)	126.5	(113.4, 139.7)
Female	70	80	81.2	(78.6, 83.7)	117.3	(112.1, 122.5)	163.1	(150.0, 176.1)
Male	50	95	97.0	(87.9, 106.1)	130.9	(120.4, 141.4)	175.6	(155.2, 196.0)
Male	70	95	117.2	(113.3, 121.1)	159.6	(151.2, 167.9)	215.1	(194.0, 236.2)
Female	50	95	116.1	(106.3, 125.9)	150.3	(135.1, 165.5)	197.3	(168.2, 226.3)
Female	70	95	122.9	(117.8: 127.9)	166.9	(156.9, 176.9)	224.7	(200.9, 248.5)

All percentiles are stated with 95% confidence intervals in parentheses. Data from OPED, 2014.

DOI: 10.1002/pds

occurs uniformly over the observation period and inter-arrival times between prescriptions conditional on covariates follow the specified distribution, Log-Normal, say, then the reverse WTD will provide statistically consistent estimates inheriting the qualities associated with maximum likelihood estimation. In contrast, the mentioned approaches are model-free and based on more or less explicit decision algorithms, which complicates general assessment of their statistical characteristics. While the reverse WTD is based on a consistent formal model, it would be of interest to assess its performance in specific settings against a gold standard measure of actual drug use. It may also be instructive to compare its performance with treatment status defined by clinical experts for a random group of patients as done by Tanskanen *et al*. This would potentially also allow assessment of which percentiles of the IAD are the most meaningful in applications.

We have developed the reverse WTD with covariates in a Danish setting where there are generally no information on dosing associated with prescription redemptions. In the Danish setting, the number of pills is a highly relevant covariate as demonstrated by our results, whereas in other settings, other covariates may be more relevant. The method is, however, generic and would allow inclusion of such covariates. Further, in settings unlike the Danish, where information on dosing is actually recorded, the information may not coincide with the actual usage patterns of patients. Then the reverse WTD with covariates may provide an opportunity to investigate discrepancies between intended and actual use. Warfarin is a medication primarily used chronically, which we considered an ideal test case for the new method. How well the method performs for other medications should be investigated, but because it is based on a generic model, its performance will be governed by the degree to which the assumptions underlying the model are satisfied.

A further advantage of the suggested method is its use of ordinary regression techniques when modelling prescription durations. Although it only allows linear combinations of covariates, this can, as in any regression model, easily be extended by inclusion of splines, quadratic terms or interactions. Linearity assumptions and effect modifications can thus be examined using standard regression techniques, which provide flexibility.

The major challenge of the method is misspecification, as for any parametric model analyzed with maximum likelihood. Misspecification may occur with respect to the parametric form of the reverse WTD, with respect to omission of important covariates, or by misspecification of the relationship between a covariate and the dependent parameter. It is therefore crucial to assess the fit of the model in diagnostic plots and by extension of the model to include splines, quadratic terms or interactions. More research on how to assess fit for this type of models is however warranted. We have in this study focused on using the Log-Normal version of the BRD (the prevalent part of the reverse WTD), because it has attractive interpretations of model parameters and was found by visual inspection to provide acceptable fit in relevant subgroups of the data. Other possibilities do, however, exist such as the simpler Exponential (one parameter: the rate) or the Weibull (two parameters: shape and scale), which in certain settings may be better choices. If so, care should be taken in interpreting effects such that they become meaningful. However, as the interpretation of the percentiles of the IAD is identical regardless of distribution choice for the BRD, the distribution-specific parameters (rate, shape and scale) may be glossed over, and one may focus on optimizing model fit when the sole aim is to predict percentiles of the IAD. These percentiles can be directly used as estimates of prescription durations, because they represent the time after a prescription within which a certain percentage of prevalent users will have returned to a pharmacy to redeem their next prescription. By definition, patients stopping treatment will have longer durations until they, possibly, re-initiate treatment, although some misclassification will remain because of random variation. The magnitude can, however, be controlled by considering different percentiles. The interpretation given here is completely analogous with our two previous papers,[10,12] except that percentiles are no longer assumed to be a 'one-size fits all' but can be tailored to individual package sizes and person characteristics. The main advantage of introducing covariates is that of any regression model, predictions are expected to become more precise, which in this context implies that misclassification of exposure status should decrease, when applying the estimated percentiles to define prescription durations. Without covariates, the 80th percentile of the IAD of warfarin is estimated at 87 days, but including information on sex, age and number of pills redeemed in the model, we found that predicted percentiles varied between 63 and 225 days for typical combinations of covariate values (Table 3). As would be expected, we found the number of pills to be the strongest predictor of prescription durations, and the ability of the method to incorporate this represents a substantial improvement over the ordinary

DOI: 10.1002/pds

WTD approach, which did not allow covariates.[18] While the precision of predictions can be expected to improve when including covariates, we cannot explicitly confirm this based on the data available to us. Consequently, future studies should examine the agreement between model-based predictions and actual drug usage among patients.

In conclusion, we have proposed an extended version of the reverse WTD, in which parameters may depend on covariates. This allows flexible modelling in which both the fraction of prevalent users and characteristics of prescription durations among prevalent users of a drug may depend on observed characteristics of the patient (age and sex, say) and the redemption (number of pills, say). This facilitates automated construction of detailed decision rules for prescription durations, which will be more precise than a single decision rule common for all patients. The method may either circumvent the need for or supplement *ad hoc* decision rules based on expert knowledge, as it provides estimates based solely on observed data. Future work should explore how better model diagnostic tools can be developed and if predictions that are more precise can be obtained by inclusion of more prescription redemptions than just the last one of each patient.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

KEY POINTS

- Many pharmacoepidemiological databases do not record information on prescription duration.
- The reverse waiting time distribution considers the last prescription of each patient before an index time point
- Using the reverse waiting time distribution, maximum likelihood estimates of regression coefficients for the prevalence fraction and the inter-arrival distribution can be obtained together with uncertainty intervals.
- Estimates of prescription durations accounting for patient and redemption characteristics can be obtained.
- Estimates can be found in STATA using the publicly available -wtdttt- package.

ETHICS STATEMENT

Approval from an Ethics Committee was not required according to Danish law.[19]

ACKNOWLEDGEMENT

None.

PREVIOUS PRESENTATIONS

This study has not previously been presented.

REFERENCES

1. Mantel-Teeuwisse AK, Klungel OH, Verschuren WMM, Porsius A, de Boer A, de Boer A. Comparison of different methods to estimate prevalence of drug use by using pharmacy records. *J Clin Epidemiol* 2001; **54**(11): 1181–1186. https://doi.org/10.1016/S0895-4356(01)00396-1.
2. Støvring H, Andersen M, Beck-Nielsen H, *et al.* Counting drugs to understand the disease: the case of measuring the diabetes epidemic. *Popul Health Metrics* 2007; **5**(1): 2. https://doi.org/10.1186/1478-7954-5-2.
3. Gardarsdottir H, Souverein PC, Egberts TCG, Heerdink ER. Construction of drug treatment episodes from drug-dispensing histories is influenced by the gap length. *J Clin Epidemiol* 2010; **63**(4): 422–427. https://doi.org/10.1016/j.jclinepi.2009.07.001.
4. Nielsen LH, Løkkegaard E, Andreasen AH, Hundrup YA, Keiding N. Estimating the effect of current, previous and never use of drugs in studies based on prescription registries. *Pharmacoepidemiol Drug Saf* 2009; **18**(2): 147–153. https://doi.org/10.1002/pds.1693.
5. Abrahamowicz M, Beauchamp M-E, Sylvestre M-P. Comparison of alternative models for linking drug exposure with adverse effects. *Stat Med* 2011. https://doi.org/10.1002/sim.4343.
6. Tanskanen A, Taipale H, Koponen M, *et al.* From prescriptions to drug use periods – things to notice. *BMC Res Notes* 2014; **7**(1): 796. https://doi.org/10.1186/1756-0500-7-796.
7. van Gaalen RD, Abrahamowicz M, Buckeridge DL. The impact of exposure model misspecification on signal detection in prospective pharmacovigilance. *Pharmacoepidemiol Drug Saf* 2015; **24**(5): 456–467. https://doi.org/10.1002/pds.3700.
8. 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. https://doi.org/10.1002/pds.3459.
9. Støvring H, Pottegård A, Hallas J. Determining prescription durations based on the parametric waiting time distribution. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1451–1459. https://doi.org/10.1002/pds.4114.
10. Støvring H, Pottegård A, Hallas J. Estimating medication stopping rates and real-time prevalence in pharmacoepidemiologic databases using the reverse waiting time distribution. *Pharmacoepidemiol Drug Saf* 2017. https://doi.org/10.1002/pds.4217.
11. Støvring H, Vach W. Estimation of prevalence and incidence based on occurrence of health-related events. *Stat Med* 2005; **24**(20): 3139–3154. https://doi.org/10.1002/sim.2142.
12. Støvring H, Pottegård A, Hallas J. Determining prescription durations based on the parametric waiting time distribution. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1451–1459.
13. Gaist D, Sorensen HTT, Hallas J. The Danish prescription registries. *Dan Med Bull* 1997; **44**(4): 445–448.
14. StataCorp. Stata Statistical Software: Release 14. 2015.
15. Meid AD, Heider D, Adler J-B, *et al.* Comparative evaluation of methods approximating drug prescription durations in claims data: modeling, simulation, and application to real data. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1434–1442. https://doi.org/10.1002/pds.4091.
16. Tanskanen A, Taipale H, Koponen M, *et al.* From prescription drug purchases to drug use periods – a second generation method (PRE2DUP). *BMC Med Inform Decis Mak* 2015; **15**(1): 21. https://doi.org/10.1186/s12911-015-0140-z.

17. Taipale H, Tanskanen A, Koponen M, Tolppanen A-M, Tiihonen J, Hartikainen S. Agreement between PRE2DUP register data modeling method and comprehensive drug use interview among older persons. *Clin Epidemiol* 2016; **8**: 363–371. https://doi.org/10.2147/CLEP.S116160.
18. Støvring H, Pottegård A, Hallas J. Determining prescription durations based on the parametric waiting time distribution. *Pharmacoepidemiol Drug Saf* 2016. https://doi.org/10.1002/pds.4114.
19. Thygesen LC, Daasnes C, Thaulow I, Bronnum-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. https://doi.org/10.1177/1403494811399956.

SUPPORTING INFORMATION

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

DOI: 10.1002/pds