PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2016; **25**: 1451–1459
Published online 26 September 2016 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.4114

ORIGINAL REPORT

Determining prescription durations based on the parametric waiting time distribution

Henrik Støvring[1]*, Anton Pottegård[2,3] and Jesper Hallas[2]

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

ABSTRACT

Purpose The purpose of the study is to develop a method to estimate the duration of single prescriptions in pharmacoepidemiological studies when the single prescription duration is not available.

Methods We developed an estimation algorithm based on maximum likelihood estimation of a parametric two-component mixture model for the waiting time distribution (WTD). The distribution component for prevalent users estimates the forward recurrence density (FRD), which is related to the distribution of time between subsequent prescription redemptions, the inter-arrival density (IAD), for users in continued treatment. We exploited this to estimate percentiles of the IAD by inversion of the estimated FRD and defined the duration of a prescription as the time within which 80% of current users will have presented themselves again. Statistical properties were examined in simulation studies, and the method was applied to empirical data for four model drugs: non-steroidal anti-inflammatory drugs (NSAIDs), warfarin, bendroflumethiazide, and levothyroxine.

Results Simulation studies found negligible bias when the data-generating model for the IAD coincided with the FRD used in the WTD estimation (Log-Normal). When the IAD consisted of a mixture of two Log-Normal distributions, but was analyzed with a single Log-Normal distribution, relative bias did not exceed 9%. Using a Log-Normal FRD, we estimated prescription durations of 117, 91, 137, and 118 days for NSAIDs, warfarin, bendroflumethiazide, and levothyroxine, respectively. Similar results were found with a Weibull FRD.

Conclusions The algorithm allows valid estimation of single prescription durations, especially when the WTD reliably separates current users from incident users, and may replace ad-hoc decision rules in automated implementations. Copyright © 2016 John Wiley & Sons, Ltd.

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

Received 21 March 2016; Revised 8 August 2016; Accepted 6 September 2016

INTRODUCTION

A key question in pharmacoepidemiological analyses is how to define the exposure duration that should be assigned to a prescription. While in some databases, the single prescription duration is recorded directly; this is lacking in many data sources. When using such data sources, it is necessary to decide on the duration that should be assigned to single prescriptions. Even when duration is recorded, it may not coincide with the actual use pattern of the drug, and it will be useful to have a method for assigning treatment durations, if only for validation purposes. While decision rules can be based on available clinical insights for some medications and patient groups, they will often have to be developed from observed prescription renewals. The waiting time distribution (WTD) has been suggested as a tool in this context.[1,2] The WTD is a charting of how long one has to wait before a user redeems his first prescription within a given time window. Its primary advantage is that it separates all users of a drug into two categories, those in current active use, that is, prevalent users, and those who initiate use, that is, incident users.[1,3] Pottegård and Hallas suggested defining a decision rule for the duration of a prescription based on a cutoff for when a certain percentage of prevalent users had redeemed their first prescription after the start of the observation window.[2]

As an alternative to the WTD, one can choose a cutoff for the inter-arrival density (IAD). This function

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

describes the distribution of distances (e.g., in days) between two subsequent prescriptions for a given drug among all its prevalent users. Although the two distributions (WTD and IAD) are related, they are markedly different. The prevalent component of the WTD is the so-called forward recurrence density (FRD), which is a mathematically known transformation of the IAD. As the incident component of the WTD is more or less uniform, the WTD has its modus at the beginning of the observation window, where the FRD also has its modus. In contrast, the IAD will in realistic scenarios have its mode sometime after the preceding prescription, that is, the next prescription redemption is more likely to occur sometime later rather than immediately after the preceding redemption. When placed on the same timescale, the IAD is thus right-shifted relative to the WTD, and any cut-point defined by a percentile is higher for the IAD than for the WTD. Intuitively, the IAD may seem as a more appropriate tool to define duration of exposure for a single prescription, because it directly shows the probability of a new prescription appearing, as a function of time.

We first provide basic formulas based on renewal process theory from which we develop a new algorithm for defining duration of a prescription based on a parametric WTD model. We then describe how to transform a parametric WTD into a parametric IAD, which makes the FRD component of the WTD corresponds to the same underlying renewal processes as the IAD. Next, we provide simulation results to explore the magnitude of bias associated with the parametric WTD model. We specifically investigate how the simple parametric WTD performs when there are two different subpopulations of users for the same drug, or when users switch between two different package sizes. Finally, we apply the parametric IAD distribution to the specific problem of defining duration of a single prescription in the same four model drugs as studied by Pottegård and Hallas: bendroflumethiazide, warfarin, levothyroxine, and non-steroidal anti-inflammatory drugs (NSAIDs).[2]

METHODS

For illustrative purposes, let us first consider a single patient with prescription redemptions at T_1, T_2, … . Let $D_i = T_{i+1} - T_i$ be time from prescription redemption i to the next $i+1$ redemption, and assume that D_i follows the continuous distribution F with density function f (the so-called inter-arrival distribution) and mean M. Further, we assume that all D_is are independent and identically distributed. This implies that the sequence of T_is form a renewal process. Let us next assume that we intercept this renewal process at some random time point t_0 at a time where the process has stabilized, that is, that the starting point is so far into the distant past that it does not matter for the current process. Alternatively, we can assume that the patient initiated treatment at a random point in time following a constant rate. The time from the interception point t_0 to the next redemption is then a so-called forward recurrence time $R = \min(T_i | T_i > t_0) - t_0$ with a distribution characterized by the following density function

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

The shape of this density is a consequence of length biased sampling (i.e., that longer intervals between redemptions have a proportionately higher probability of being sampled) and the interception point being uniformly distributed on intercepted intervals.

When we consider the inter-arrival distribution, F, we can define the time point τ_k for which there is $k\%$ chance of a patient redeeming a new prescription before that time. Mathematically, τ_k will then have to satisfy the following equation:

$$\int_0^{\tau_k} f(s)ds = k\%$$

Note that τ_k is not in general equal to the time point in the forward recurrence distribution at which there is $k\%$ chance of observing a smaller forward recurrence time, that is, the τ_k' defined from the following equation is not in general equal to τ_k:

$$\int_0^{\tau_k'} \frac{1 - F(r)}{M} dr = k\%$$

where τ_k' is the percentile considered by Pottegård and Hallas.[2] The two percentiles, τ_k and τ_k', are only equal when F is an exponential distribution with a constant rate. The exponential distribution has identical IAD and FRD because of its memoryless property, which implies that the subsequent redemption should occur with a constant rate, that is, regardless of the time since last prescription redemption. This is a rather unrealistic assumption for most processes of actual prescription redemptions.

To illustrate the difference between the two percentiles of the IAD and the FRD, respectively, we may consider a Log-Normal inter-arrival distribution with a mean of 90 days and a SD of 30 days. The parameters

DOI: 10.1002/pds

of the Log-Normal distribution are related to the mean, M, and variance, V, of the IAD by the formulas[4]:

$$M = \exp\left(\mu + \frac{\sigma^2}{2}\right), \text{ and } V = (e^{\sigma 2} - 1)\exp(2\mu + \sigma^2)$$

By inversion of these formulas, the log-mean of the distribution is given by $\log(90^2/\sqrt{(30^2+90^2)}) = 4.447$ and the variance on the log-scale by $\log\left(1 + \frac{30^2}{90^2}\right) = 0.105$. We can find the 80th percentile of this distribution by looking up the value in a standard normal distribution and re-transform. This yields us a value of the 80th percentile for the IAD of 112.1 days, that is, 80% of all prevalent users will have a new prescription within 112.1 days of their previous redemption. If instead we were to base it on the percentile of the corresponding forward recurrence distribution, as suggested by Pottegård and Hallas,[2] we would obtain the value 78.2 days, compare with Figures 1 and 2.

From the formulas presented earlier, it is evident that we can actually obtain the percentile τ_k for the inter-arrival distribution F from the FRD $g(t)$. Further, note that the density $g(t)$ is the component associated with prevalent users, which is directly estimated in the parametric version of the WTD approach.[3] From the density $g(t)$, we can find τ_k as the time t that fulfils the following equality:

$$M \cdot g(t) = 100\% - k\%$$

Solving this equation can be carried out either analytically or numerically after estimation of $g(t)$, and by noting that with an estimate of the parameters of the density $g(t)$, M is implicitly defined.

The algorithm we propose for estimating τ_k therefore consists of three steps:

(1) Obtain a parametric estimate of the WTD, specifically the parameters of the distribution component $g(t)$, which corresponds to prevalent users.
(2) Estimate M from the parameters of g.
(3) Find the time τ_k as the time t, such that $M \cdot g(t) = (100\% - k\%)$.

To mimic the setting of the original paper by Pottegård and Hallas[2] as closely as possible, we used a version of the parametric WTD, which does not account for censoring, that is, it only considers observed prescription redemptions. With a uniform density for the incident component of the WTD, the likelihood contribution for a single individual is then given by

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

where γ is the fraction of prevalent users among the observed users in the observation window, δ is the width of the observation window, and $g(t;\theta)$ is the FRD for prevalent users, which depends on parameters θ. In this paper, we consider two parametric distributions, Log-Normal and Weibull, which are parameterized as follows:

1. Log-Normal FRD: $g(t) = \frac{1}{M}\Phi\left(\frac{\log t - \mu}{\sigma}\right)$ where Φ is the cumulative standard normal distribution function and M is the inter-arrival mean given by $M = \exp\left(\mu + \frac{\sigma^2}{2}\right)$. The kth percentile of the inter-arrival distribution is given by $\tau_k = \exp(\Phi^{-1}(k) \cdot \sigma + \mu)$
2. Weibull FRD: $g(t) = \frac{1}{M}\exp(-(\beta \cdot t)^{\alpha})$ where M is the corresponding inter-arrival mean given by

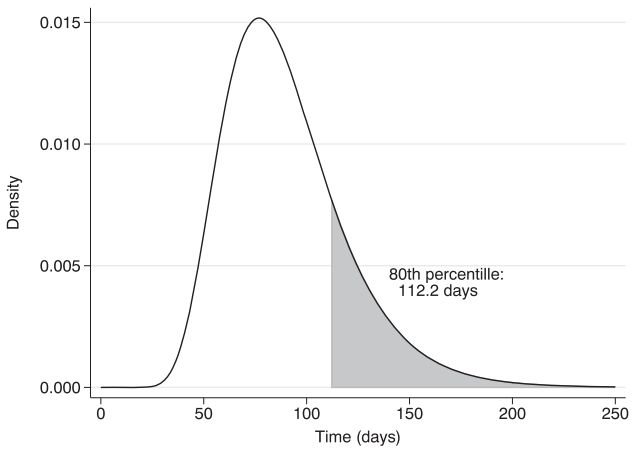

Figure 1. Density for a Log-Normal inter-arrival distribution with a mean of 90 days and a standard deviation of 30 days

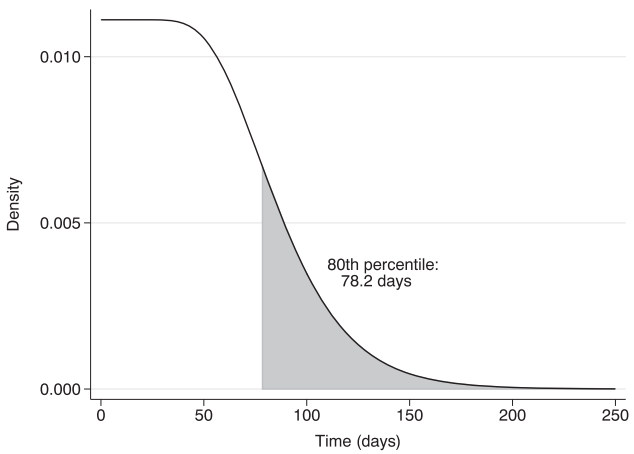

Figure 2. Forward recurrence density for a Log-Normal inter-arrival distribution with a mean of 90 days and a standard deviation of 30 days

DOI: 10.1002/pds

$M = \frac{1}{\beta}\Gamma\left(1+\frac{1}{\alpha}\right)$ and where Γ is the Gamma function defined by $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx$. The kth percentile of the inter-arrival distribution is given by $\tau_k = -\frac{1}{\beta}\sqrt[\alpha]{\log(1-k)}$

To improve convergence and stability of the maximum likelihood estimation procedure, we log-transformed the parameters σ, α, and β, as they are required to be larger than zero by definition. We logit-transformed γ. This is similar to a previous implementation of estimation for the parametric WTD.[3]

SIMULATION STUDIES

Scenario 0: Single inter-arrival density—one type of patients

Imagine a situation where all patients are treated with the same package size and where they all have the same distribution of time from one prescription redemption to the next.

Scenario 1: Two-component mixtures—two types of patients

Imagine a situation in which patients are treated with a medication, but they consist of two groups, for example, due to two different package sizes: One group of patients redeems 30 pills a time, the other group 90 pills. With an average dose of 1 pill per day, we would expect redemptions to occur either 30 or 90 days apart for each of the two groups. If we assume that patients are of either one or the other kind without changing status over the course of treatment, then the density of a randomly selected patients' time from one prescription redemption to the next is given by

$$\tilde{f}(t) = p_1 \cdot f_1(t) + (1-p_1) f_2(t)$$

Here, $f_1(t)$ is the density of the first group, $f_2(t)$ for the other group, and p_1 is the prevalence of the first type of users. Note, however, that this is different from the marginal IAD after a randomly selected prescription, because patients redeeming smaller packages will contribute more redemption over the same amount of time and will therefore have a higher chance of being sampled. The FRD is in this situation given by

$$g(t) = p_1 \frac{1-F_1(t)}{M_1} + (1-p_1)\frac{1-F_2(t)}{M_2}$$

where M_1 is the mean of the distribution F_1, that is, 30 days, and M_2 is the mean of F_2, that is, 90 days. From this FRD, we can obtain the following formula for the marginal IAD as

$$f(t) = \lambda \cdot f_1(t) + (1-\lambda) f_2(t)$$

where

$$\lambda = \frac{p_1 \cdot M_2}{p_1 \cdot M_2 + (1-p_1)M_1}$$

From this, it can be seen that if $M_1 < M_2$ then $\lambda > p_1$, that is, a higher representation of shorter intervals in the IAD. In the simulation studies, we used $p_1 = 0.7$.

Scenario 2: Two-component mixtures—one type of patients

Here, we imagine a situation where patients are treated with a medication in two different package sizes and that each time a prescription redemption occurs there is probability π_1 of redeeming a package of size 30, that is, the density for time to next prescription is $f_1(t)$, and probability $(1-\pi_1)$ for redeeming a package of size 90 (density for time to next prescription is $f_2(t)$). Then the density of the time from one prescription to the next is given by

$$f(t) = \pi_1 \cdot f_1(t) + (1-\pi_1) f_2(t)$$

In this scenario, the FRD is given by

$g(t) = \frac{\pi_1(1-F_1(t)) + (1-\pi_1)(1-F_2(t))}{\pi_1 \cdot M_1 + (1-\pi_1)M_2}$ which may be rewritten as

$$g(t) = \pi_1^* \frac{1-F_1(t)}{M_1} + \left(1-\pi_1^*\right)\frac{1-F_2(t)}{M_2}$$

where $\pi_1^* = \pi_1 \cdot M_1 / (\pi_1 \cdot M_1 + (1-\pi_1)M_2)$. In other words, the mixing probability in the forward recurrence distribution is weighted by the mean inter-arrival time M_1 of the first component, F_1, relative to the overall mean of the marginal inter-arrival distribution. In the simulation studies, we used $\pi_1 = 0.7$.

Setup for simulation studies

In all scenarios, we assume that treatment onset occurs with a constant rate and that the WTD on average consists of 25% incident users and 75% prevalent users within a 1-year observation window, that is, the true value of γ was set to 0.75. The FRD of prevalent users is assumed to either correspond to a single Log-Normal distribution for the inter-arrival times (Scenario 0), a branching process of two patient groups each with a

DOI: 10.1002/pds

Log-Normal distribution for the inter-arrival times (Scenario 1), or a single type of patients with a two-component mixture density for inter-arrival times, where each component follows a Log-Normal distribution (Scenario 2). The first component distribution has a median value of 1.5 months (45 days) in all settings, except when used on its own in Scenario 0, where in addition we consider a median value of 3 months (90 days). The second component distribution either has a median of 2 (60 days) or 3 months (90 days). The variation of the first component distribution is always such that its central 95% range is a factor of 1.5 on either side of the median, that is, the 97.5% percentile of the density is 1.5 times the median. For the second component, the corresponding variation factor is either 1.5 or 2. An overview of the distributions can be seen in Figure 2a, b, and c in the Supporting Information. It should be noted that despite the difference in setup between scenario 1 and 2, the type of the generated WTD will be the same—indeed, adequately chosen parameters for scenario 1 will produce a WTD identical to that of scenario 2, and vice versa.

We study each simulation setting with three different sample sizes ($n = 1,000$; $n = 5,000$; or $n = 20,000$), where n is the number of patients with a prescription redemption. For each setting, we generate 2,500 datasets. We used the algorithm proposed by Zhao to generate samples from Log-Normal FRDs.[5] The parameters of the Log-Normal distribution were derived from the specified median and variation factor as follows: $\mu = \log(\text{Median})$, and $\sigma = \log(\text{Variation factor})/\Phi^{-1}(.975)$

All generated datasets are subsequently analyzed using a parametric WTD approach with a uniform density for the incidence component and a Log-Normal FRD for the prevalent component. For each setting, we estimate the relative bias in estimates of $\tau_{80\,\%}$, its empirical standard deviation and the root mean square error (RMSE) around the true value.

APPLICATION

We analyze the same four model drugs as Pottegård and Hallas: NSAIDS, warfarin bendroflumethiazide and levothyroxine.[2] For each of the four drugs, we analyze two samples: The first sample consists of all the first prescriptions of users in 2009 redeemed in the Region of Southern Denmark (1.2 million inhabitants) and thereby captured in Odense Pharmacoepidemiological Database.[6] In the second sample, we only consider the first prescription of a user in 2009, if the same user also redeemed a prescription for the same type of drug in 2008, because this restriction was suggested by Pottegård and Hallas.[2] The indication, dosage, and refill instruction are not recorded in OPED, and thus, the duration of a prescription 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.

All statistical analyses were conducted in Stata 14.1.[7] A dedicated software package (wtd_perc) 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 wtd_perc, all–.

RESULTS

Simulation studies

Results of the simulation study are shown in Table 1. When the WTD was analyzed with the correct model corresponding to a single Log-Normal distribution for inter-arrival times, the relative bias was negligible (−0.20% to 0.01%) regardless of sample size and amount of variation in the inter-arrival distribution. Precision of the estimate increased with sample size and with 20,000 observations the RMSE was 0.61 days, which indicates that the 80% percentile of the inter-arrival distribution can be estimated within approximately ±1 day of the true value with 95% confidence, when the FRD is correctly specified.

For both of the other two simulation scenarios, the relative bias was smallest when the two densities governing the inter-arrival distribution had medians closer to each other (1.5 and 2 months). The largest relative bias of 8.14% was seen for the branching process (Scenario 1) where the second type of distribution had a median of 3 months and a variation factor of 2. For the two-component distribution (Scenario 2), the numerically largest relative bias of −5.93% ($n = 1000$) was seen when the second distribution had a median of 3 months and a variation factor of 1.5. Because the empirical standard deviations of the estimates were rather similar across all eight settings in Scenarios 1 and 2, variations in the magnitude of the RMSE were dominated by variation in the bias. The largest RMSEs thus occurred together with the largest relative biases but did not in any of the settings exceed 6 days. Estimates in all settings closely followed a normal distribution, which suggests that confidence intervals can be validly obtained by normal approximation and a bootstrap estimate of the standard error (graphs not shown, available from the first author upon request).

DOI: 10.1002/pds

Table 1. Simulation results for three different types of scenarios

					$n = 1000$			$n = 5000$			$n = 20\,000$		
Scenario	Median, 1st density (months)	Median, 2nd density (months)	Variation factor, 2nd density	True IAD 80% percentile (days)	Relative bias (%)	SD (days)	RMSE (days)	Relative bias (%)	SD (days)	RMSE (days)	Relative bias (%)	SD (days)	RMSE (days)
0	1.5	N/A	N/A	54.34	−0.20	1.21	1.22	0.01	0.52	0.52	−0.01	0.27	0.27
0	3	N/A	N/A	108.68	−0.08	2.73	2.74	−0.01	1.22	1.22	−0.01	0.61	0.61
1	1.5	2	1.5	59.95	0.04	1.52	1.52	0.17	0.65	0.66	0.17	0.33	0.34
1	1.5	2	2	59.54	2.46	1.80	2.32	2.47	0.81	1.68	2.51	0.41	1.55
1	1.5	3	1.5	65.54	4.82	2.40	3.96	4.95	1.07	3.42	4.89	0.54	3.25
1	1.5	3	2	62.49	8.05	2.63	5.68	8.18	1.22	5.26	8.14	0.59	5.12
2	1.5	2	1.5	61.30	−0.17	1.52	1.52	0.00	0.68	0.68	−0.01	0.35	0.35
2	1.5	2	2	61.42	2.57	1.97	2.53	2.63	0.85	1.82	2.61	0.43	1.66
2	1.5	3	1.5	83.72	−5.93	2.85	5.72	−5.92	1.31	5.13	−5.85	0.62	4.94
2	1.5	3	2	79.11	0.16	3.29	3.29	0.30	1.50	1.52	0.34	0.73	0.78

A single Log-Normal inter-arrival density (Scenario 0), a branching process where there is a 70% chance that an individual has the first Log-Normal density as their inter-arrival density, 30% the second Log-normal density (Scenario 1), or the inter-arrival density for all individuals consists of a mixture of two Log-Normal densities (Scenario 2), where 70% percent of all inter-arrival times come from the first density and 30% from the second density. All Log-Normal densities are specified by their median value and by the factor governing the width of their central 95% range relative to the median, that is, a factor of 1.5 means that the 97.5% percentile of the density is 1.5 times the median. For the first density, the variation factor was 1.5 in all scenarios. In all scenarios, the window for observing the forward recurrence times was 1 year, and we assumed that the true proportion of incident individuals was 25% and the remaining 75% were prevalent. For each sample size, 2,500 datasets of the forward recurrence times were generated.

Relative bias was computed as the average relative difference between the estimated 80% percentile and the true 80% percentile.

SD is the standard deviation observed of the estimates of the 80% percentile.

RMSE is the root mean square error of the estimates around the true value of the 80% percentile.

DOI: 10.1002/pds

Empiric studies

When the parametric WTD was applied to the 2009 data for the four model drugs in the unrestricted sample, we found 80th percentiles that were substantially larger than those previously reported by Pottegård and Hallas, except for NSAIDs (Table 2). The estimated percentiles were largely unaffected whether the FRD was taken to be Log-Normal or Weibull, again with NSAIDs as the exception.

For warfarin, bendroflumethiazide and levothyroxine the parametric WTD can be expected to reliably separate the two components of the WTD corresponding to prevalent and incident users. This is supported by the results for the restricted sample of users, where we required patients to also have redeemed a prescription in 2008. With the parametric WTD, estimated percentiles were almost identical in the restricted and unrestricted sample. This is a consequence of the observation that the restricted sample for warfarin, bendroflumethiazide, and levothyroxine may reasonably be considered a prevalent sample. For NSAIDs, the differences were more pronounced between estimates from the restricted and unrestricted sample, although still much more similar to each other than to the estimate of Pottegård and Hallas.[2]

DISCUSSION

In settings with a substantial component of prevalent use of a drug, the proposed parametric WTD can reliably estimate the time at which a given proportion of users in continued treatment will have redeemed a new prescription after their previous prescription, that is, it can be used to estimate a percentile of their inter-arrival distribution. When a substantial proportion of patients use a drug intermittently, the method becomes more dependent on choice of parametric model, and hence less robust, although it seems to provide more reliable estimates of inter-arrival distribution percentiles than the previously suggested method based on the WTD.[2]

The primary advantage of the method is that it presents an opportunity to assign duration exposure to prescriptions with an automated algorithm based on observed prescription redemption patterns. Further, the assigned duration has a straightforward interpretation based on a mathematically sound model and with valid estimates of uncertainty, that is, confidence intervals. Our simulation study showed that the estimation was unbiased when the specified model was correct, that is, when it coincided with the model used to generate the data analyzed in simulations. Even when the

Table 2. Estimated 80th percentile for prescription duration for four different model drugs: NSAIDS, warfarin, bendroflumethiazide and levothyroxine

2008 as run-in period?	Drug	Parametric model	*n*	Parametric WTD 80th percentile (days)	PH2013 80th percentile (days)*	Relative difference (%)†
No	NSAID	Log-Normal	182 909	116.5 (113.5; 119.5)	210	80.3
No	NSAID	Weibull	182 909	130.1 (127.2; 133.1)	210	61.4
No	Warfarin	Log-Normal	17 777	91.2 (89.3; 93.1)	69	−24.3
No	Warfarin	Weibull	17 777	91.4 (89.6; 93.2)	69	−24.5
No	Bendroflumethiazide	Log-Normal	3822	137.0 (130.5; 143.9)	92	−32.9
No	Bendroflumethiazide	Weibull	3822	137.5 (129.2; 146.3)	92	−33.1
No	Levothyroxine	Log-Normal	24 178	118.3 (116.6; 120.0)	86	−27.3
No	Levothyroxine	Weibull	24 178	117.7 (116.1; 119.3)	86	−26.9
Yes	NSAID	Log-Normal	70 989	111.4 (109.3; 113.5)	210	88.5
Yes	NSAID	Weibull	70 989	116.3 (114.2; 118.4)	210	80.5
Yes	Warfarin	Log-Normal	13 756	90.0 (88.3; 91.7)	69	−23.3
Yes	Warfarin	Weibull	13 756	90.2 (88.6; 91.8)	69	−23.5
Yes	Bendroflumethiazide	Log-Normal	2483	132.6 (128.0; 137.5)	92	−30.6
Yes	Bendroflumethiazide	Weibull	2483	133.1 (128.0; 138.5)	92	−30.9
Yes	Levothyroxine	Log-Normal	21 245	116.9 (115.5; 118.3)	86	−26.4
Yes	Levothyroxine	Weibull	21 245	116.5 (115.0; 118.0)	86	−26.2

WTD, waiting time distribution; NSAIDs, non-steroidal anti-inflammatory drugs.
The WTD used either a Log-Normal or a Weibull forward recurrence density for the prevalent component, whereas the incident component was modelled with a uniform distribution. When 2008 was used as run-in period, the analyzed WTD sample was restricted to users who also redeemed a prescription in 2008. Confidence intervals were obtained with a bootstrapped SE based on 50 samples.
*Estimated percentiles reported by Pottegård and Hallas.[2]
†The relative difference between the Pottegård and Hallas[2] estimate and the estimate based on the parametric WTD.

DOI: 10.1002/pds

data-generating model differed markedly from the fitted model, the relative bias was tolerable as it did not exceed 8%. If the inter-arrival distribution had more than two modes, the corresponding WTD would be smoothed further, and the scenarios with two modes may therefore be considered the most extreme case. In our application of the method to four model drugs, we found that the method robustly estimated the 80th percentile of the inter-arrival distribution for drugs with a substantial component of prevalent use, irrespective of choice of parametric model (Log-Normal or Weibull), and regardless of whether the sample was restricted to users with a prescription in the previous year or not. Our approach may thus obviate the need for a 1-year run-in period, as used by Pottegård and Hallas, which is useful when using data sources with substantial annual exchange of population.

The main weakness of the model is shared with the original approach of Pottegård and Hallas[2] as it does not allow for incorporation of individual characteristics of the patient such as past patterns in prescription redemptions or current dose redeemed. Also, it requires the WTD to be able to reliably separate users into two categories of current users, prevalent users and incident users initiating treatment, as was also the case for the method by Pottegård and Hallas.[2] It is therefore less useful for drugs with substantial intermittent use, although our finding of nearly identically estimated percentiles from the restricted and unrestricted sample may indicate that the parametric WTD is to some extent better at identifying the prevalent component of the WTD than the Pottegård and Hallas approach. The larger estimate of Pottegård and Hallas for NSAIDs may be a consequence of misclassifying users as being prevalent on 1 January 2009, whenever they had a prescription in 2008—likely many who were users at some point during 2008 will have stopped their use of NSAIDs before the end of the year. When these users then reinitiate treatment in 2009, it will appear to be from a prevalent user, and the Pottegård and Hallas method will overestimate the percentile of interest, despite its inherent theoretical tendency to underestimate. When there is substantial intermittent use, the parametric WTD approach does become more sensitive to choice of parametric distribution, as it is difficult to separate the uniform distribution for incidence from a slowly declining FRD for prevalence in the NSAID example (graph not shown, available upon request). As with all parametric methods, the fit of the model to the observed data determines the amount of bias. In the simulation studies, the highest bias was generally seen, when data were generated from with two distinctively different subdistributions of the FRD. In applications, it is therefore important to inspect visually the empirical WTD to see if there is a clear uniform part towards the end of the observation window and a smoothly declining part in the beginning. This follows suggestions based on previous simulation studies for the parametric WTD.[3]

Studies have shown that choice of prescription duration influences apparent duration of treatment episodes and estimated risk associated with exposure. McMahon *et al*. showed how different lengths of exposure periods assigned to NSAID prescriptions led to relative risk estimates for gastrointestinal hemorrhage varying from 2.16 to 5.82.[8] While they argued for not letting exposure periods exceed prescription durations, they did not provide guidance for choosing adequate prescription durations. Gardarsdottir *et al*. showed that varying the length of prescription durations for selective serotonin reuptake inhibitors could double the median antidepressant treatment episode length.[9] No optimal approach has, however, been identified, and consequently, sensitivity analyses remain the recommended approach because it allows exploration of the robustness of results with respect to varying prescription durations.[10,11]

When the parametric WTD is used to estimate a percentile of the IAD among continued users, it may be viewed as putting an upper limit on the misclassification of continued users: With a prescription duration defined from the 80th percentile, 20% of continued users will mistakenly be classified as having stopped use. Such considerations may help inform the choice of which percentile to estimate, because different levels of misclassification may be desired in different applications. Potentially, the choice of a large percentile, say 95%, may obviate the need for grace periods, although at the cost of increasing misclassification of stopped users as having continued use. The amount of misclassification of stopped users from such an approach is not known and should be investigated in future studies.

As the method is based on a parametric model and maximum likelihood estimation, it should in principle be possible to extend it with incorporation of covariates of interest. In many applications, information on package size is available, and likely, this will be informative about the duration to the next prescription redemption. The package size of prescriptions considered in the WTD information will, however, not be directly informative of the WTD, because it will be the size of the previous prescription redemption that affects the shape of the WTD. How to incorporate this merits further research but is beyond the scope of the

DOI: 10.1002/pds

present paper. The estimate provided here is a marginal population estimate, but estimates may be obtained for subgroups defined by gender and/or age categories. Further, as the method provides an uncertainty estimate of the estimated percentile, it should be investigated how this could be incorporated in subsequent analyses relying on the estimated value of the percentile.

In conclusion, we suggest that the parametric WTD may be used to estimate percentiles of the distribution of time between subsequent prescription redemptions for continued use of a drug and that this can be used in an automated fashion to assign exposure duration to prescriptions in pharmacoepidemiological research. Future studies should consider how to improve robustness of the method by incorporating the observed distribution time between subsequent prescription redemptions and information on characteristics of the patient and the redeemed prescription.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PREVIOUS PRESENTATIONS

This study has not previously been presented.

KEY POINTS

- Many pharmacoepidemiological databases do not record information on prescription duration
- No optimal decision rules exist for determining prescription durations.
- Maximum likelihood estimation of the waiting time distribution allows reliable estimation of the 80th percentile of the inter-arrival distribution (time between subsequent prescription redemptions) in situations where most users are prevalent users.
- The method does not require information on redemptions in a preceding period and may be automated.
- A package (wtd_perc) for estimation within Stata is available for download.

ETHICS STATEMENT

Approval from an Ethics Committee was not required.

REFERENCES

1. Hallas J, Gaist D, Bjerrum L. The waiting time distribution as a graphical approach to epidemiologic measures of drug utilization. *Epidemiology* 1997; **8**: 666–70.
2. Pottegård A, Hallas J. Assigning exposure duration to single prescriptions by use of the waiting time distribution. *Pharmacoepidemiol. Drug Saf.* 2013; **22**: 803–809. doi:10.1002/pds.3459.
3. Stovring H, Vach W. Estimation of prevalence and incidence based on occurrence of health-related events. *Stat. Med.* 2005; **24**: 3139–54.
4. Wiley: *Continuous Univariate Distributions*, **1**, 2nd Edition Norman L. Johnson, Samuel Kotz, N. Balakrishnan. Available at: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471584959.html. Accessed July 5, 2016.
5. Zhao Y. Parametric inference from window censored renewal process data. 2006. Available at: https://etd.ohiolink.edu/rws_etd/document/get/osu1164678679/inline.
6. Gaist D, Sorensen HT, Hallas J. The Danish prescription registries. *Dan. Med. Bull.* 1997; **44**: 445–8.
7. StataCorp. *Stata Statistical Software: Release 14*. StataCorp LP: College Station, TX, 2015.
8. McMahon AD, Evans JMM, McGilchrist MM, McDevitt DG, MacDonald TM. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. *Pharmacoepidemiol. Drug Saf.* 1998; **7**: 275–280.
9. 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**: 422–427. doi:10.1016/j.jclinepi.2009.07.001.
10. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J. Clin. Epidemiol.* 2005; **58**: 323–337. doi:10.1016/j.jclinepi.2004.10.012.
11. Hallas J, Stovring H. Templates for analysis of individual-level prescription data. *Basic Clin. Pharmacol. Toxicol.* 2006; **98**: 260–265. doi:10.1111/j.1742-7843.2006.pto_257.x.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.