Received: 19 January 2021 | Revised: 21 May 2021 | Accepted: 9 August 2021

DOI: 10.1002/pds.5340

ORIGINAL ARTICLE

WILEY

Using multiple random index dates with the reverse waiting time distribution improves precision of estimated prescription durations

Katrine Bødkergaard[1,2] | Randi Marie Selmer[3] | Jesper Hallas[4] | Lars Jøran Kjerpeseth[3] | Eva Skovlund[3,5] | Henrik Støvring[1]

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

[2]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark

[3]Chronic Diseases and Ageing, Norwegian Institute of Public Health, Oslo, Norway

[4]Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, Odense, Denmark

[5]Department of Public Health and Nursing, NTNU, Trondheim, Norway

Correspondence
Katrine Bødkergaard, Department Clinical Epidemiology, Aarhus University, Olof Palmes Allé 43-45, 8200 Aarhus N, Denmark.
Email: kani@clin.au.dk

Abstract

Purpose: To improve the precision of prescription duration estimates when using the reverse waiting time distribution (rWTD).

Methods: For each patient we uniformly sampled multiple random index dates within a sampling window of length δ. For each index date, we identified the last preceding prescription redemption, if any, within distance δ. Based on all pairs of last prescription and index date, we estimated prescription durations using the rWTD with robust variance estimation. In simulation studies with increasing misspecification we investigated bias, root mean square error (RMSE) and coverage probability of the rWTD using multiple index dates (1, 5, 10, and 20). We applied the method to Danish data on warfarin prescriptions from 2013 to 2014 stratifying by and adjusting for sex and age.

Results: In simulation scenarios without misspecification, the relative bias was negligible (−0.04% to 0.01%) and nominal coverage probabilities almost retained (93.8%–95.4%). RMSE decreased with the number of random index dates (e.g., from 1.3 with 1 index date to 0.6 days with 5). With misspecification, the relative bias was higher irrespective of the number of index dates. Precision increased with the number of index dates, and hence coverage probabilities decreased. When estimating durations of warfarin prescriptions in Denmark, precision increased with number of index dates, in particular in strata with few patients (e.g., men 90+ years: width of 95% confidence interval was 16.2 days with 5 index dates versus 35.4 with 1).

Conclusions: Increasing the number of random index dates used with the rWTD improved precision without affecting bias.

KEYWORDS

maximum likelihood, parametric modeling, pharmacoepidemiology, precision, prescription duration, waiting time distribution

Key Points

- Many pharmacoepidemiological databases do not record information on prescription duration.

Previous presentations: This work has been presented as a poster at the virtual conference, ICPE 2020 All Access.

- The reverse Waiting Time Distribution (rWTD) considers the last prescription of each patient before an index time point.
- The rWTD with only one index date (fixed or random) is statistically inefficient.
- Using the rWTD with multiple random index dates improves precision.
- Estimates can be found in Stata using the publicly available -wtdttt- package.

1 | INTRODUCTION

How to determine prescription durations and hence exposure status remains an important question in most pharmacoepidemiological studies based on register data of redeemed prescriptions. Often, studies rely on simple decision rules based on assumed consumption of a defined daily dose or one tablet per day. However, not all prescription registries contain information on the prescribed daily dose or days' supply, and even when they do the information may not be consistent with actual intake. We have therefore suggested the parametric Waiting Time Distribution (WTD) as a flexible model-based alternative, which provides estimates of prescription durations only relying on observed prescription data using a formal statistical model.[1–3] The WTD is an umbrella term for both the ordinary and reverse WTD (rWTD). The methods rely on selecting a fixed index date and then consider either the time to the first subsequent redemption (ordinary) or from the last previous redemption (reverse) for each patient, in both cases within a fixed time window. The WTD has been developed for settings, where drugs are dispensed repeatedly to each patient and it assumes an absence of seasonality in rates of prescription redemptions and treatment stopping. However, even in situations with detectable misspecification, the WTD has been shown to provide estimates with modest bias.[1] Further, in a recent extension of the model, we have introduced random sampling of index dates to mitigate the effect of seasonal variation in rates of prescription redemptions.[4] In settings with seasonal variation, estimated prescription durations will vary markedly with the location of a single fixed index date, whereas using individual random index dates will provide a stable marginal estimate of the prescription duration for the entire sampling period.

Current versions of the WTD all use a single index date for each patient, which is statistically inefficient as it discards most of the observed data. Especially when data is only available for few patients or prescription durations need to be estimated in subgroups or adjusted for multiple covariates, estimates may have low precision. Building upon our previous approach of using a single random index date, we suggest to improve the precision by sampling multiple random index dates for each patient. By doing so, information from more redemptions are used for each patient, which increases the effective sample size. To account for the dependence of observations originating from the same patient, we use robust variance estimation. The focus of this paper is on the rWTD, since it allows covariates, but the extension with sampling multiple random index dates is also applicable for the ordinary case, although then without covariates.

We first describe the modification of the rWTD with random index dates to sampling more index dates for each patient within a pre-specified sampling window. We then investigate the performance of the method in simulation studies both without and with seasonal variation in rates of prescription redemptions. Finally, we apply the method to Danish data on warfarin redemptions.

2 | METHODS

The rWTD using either a fixed index date for all, t_0, or individual random index dates, t_{0i}, has been introduced in previous papers.[2,4] Briefly, the rWTD is a parametric two-component model consisting of a prevalent component, given by the backward recurrence density (BRD), and a uniform stopping component (Appendix A, Figure A1). The BRD is the distribution of the time from the last preceding redemption to a given index time, r, and is related to the inter-arrival distribution (IAD), which governs the time between redemptions for patients continuing treatment. If we let the distribution function F denote the IAD depending on parameters θ and with mean M, it is related to the density function g of the BRD through the formula

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

For the approach with a fixed index date only redemptions within the observation window $(t_0 - \delta; t_0)$ are included in the rWTD analysis, whereas with random index dates the observation windows are individual $(t_{0i} - \delta; t_{0i})$. The individual random index dates are uniformly sampled from a sampling window of length δ, typically a calendar year in applications. Hence, to perform a rWTD analysis with random index dates, a data window of total length 2δ is needed so as to contain all individual observation windows.

We now consider sampling multiple random index dates for each patient, t_{0ij}, within the sampling window (see Figure 1). Here t_{0ij} denotes the j'th random index date for patient i.

For each patient and each index date, we consider the observation window $(t_{0ij} - \delta; t_{0ij})$ and identify the last prescription redemption within this observation window, if any (see Figure 2).

The time r_{ij} is the time from this last redemption to the index date t_{0ij}. As we base the analysis on the r_{ij}'s we have effectively shifted the time scales for each patient and each index date to align all index dates. We can therefore include all the times from the pairs of last preceding redemption until the given index date in the rWTD analysis

in the same manner as for the analysis with one random index time for each patient.[4] Since we have more observations from the same patient, we use robust variance estimation to account for this non-independence in the data. Similar to the previously introduced WTD methods, we estimate prescription durations as the time within which 80% of the prevalent patients have redeemed a new prescription, $\tau_{80\%}$, i.e. the 80th percentile of the IAD.[1–4]

The rWTD depends on the parameters θ, from the IAD, and p, the fraction of prevalent users among the observed users in the observation window. As for the rWTD with a single fixed index date the parameters of the rWTD can depend on covariates.[3]

2.1 | Simulation studies

We evaluated the performance of the proposed methods in four scenarios with varying degrees of misspecification, as described below—for further details see the Appendix A. To allow transparent comparison, the considered scenarios are the same as those used in a previous study to examine the properties of the WTD with a single random index date.[4]

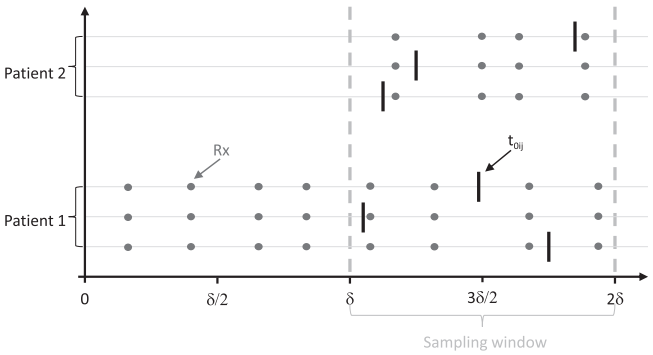

FIGURE 1 Prescription redemption dates, Rx, for two individuals. For each patient, the drug purchase history is replicated three times corresponding to three different random index dates, t_{0ij}, sampled uniformly within the sampling window

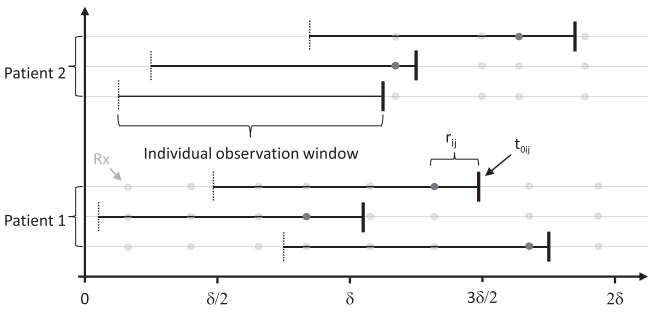

FIGURE 2 Prescription redemption dates, Rx, for two individuals. For each patient, the drug purchase history is replicated three times corresponding to three different random index dates, t_{0ij}. The last observation (dark gray) in the individual random observation windows, $(t_{0ij} - \delta; t_{0ij})$, are identified and we consider the time from the index date to the last observation, r_{ij}

All scenarios assume that patients have a treatment episode consisting of a sequence of prescription durations. The prevalence of treated patients is assumed constant over time due to a constant incidence rate and a stable distribution of the length of treatment episodes. The durations of treatment episodes vary independently of the distribution of prescription durations within the treatment episode as described below. For convenience in the simulations, stockpiling can occur in the same way for both prevalent and incident users initiating treatment in the stockpiling period. The scenarios are briefly described below, for further details see Appendix A.

- **Scenario 1**: No stockpiling and constant probability that any prescription redemption is the last of a treatment episode—no misspecification.
- **Scenario 2**: No stockpiling and Log-Normally distributed durations of treatment episodes—misspecification with respect to the treatment stopping process.
- **Scenario 3**: Stockpiling due to larger redemptions at the end of the year and Log-Normally distributed durations of treatment episodes—misspecification with respect to the homogeneity of a single Log-Normal IAD.
- **Scenario 4**: Stockpiling due to more frequent redemptions at the end of the year and Log-Normally distributed durations of treatment episodes—again misspecification with respect to the IAD homogeneity as in Scenario 3.

Let IAD_1 and IAD_2 denote the Log-Normal IADs for prescriptions redeemed in periods without and with stockpiling, respectively. For Scenario 4 we further have IAD_3, which denotes the Log-Normal IAD for the last prescription of the year redeemed after stockpiling.

The simulated datasets are analyzed using the rWTD with 1, 5, 10, and 20 random index dates for each patient, respectively. The prevalent component of the rWTD corresponds to a single Log-Normal IAD. We estimate the relative bias, root mean square error (RMSE) and coverage probability of nominal 95% confidence intervals with respect to $\tau_{80\%}$ for each setting. By definition, the RMSE can be decomposed into a bias and standard error term, where the latter is a measure of precision.

The Stata code for the simulation scenarios are available upon request.

2.2 | Application

In the empirical analysis we used Danish data on prescription redemptions of warfarin in the years 2013 and 2014.[5] We applied the new method with 1, 5, 10, and 20 random index dates for each patient, respectively, and a Log-Normal prevalent component with parameters μ and σ. We conducted two analyses; one stratified by sex and age group (0–49, 50–59, 60–69, 70–79, 80–89, and 90+ years); and a joint model for both sexes and all age groups where the parameters of the rWTD (p, μ, and σ) depend on sex and age group. For details on the model likelihood see Appendix B, which also shows results from a model exploring how parameters in this case depends on sex and age group.

All statistical analyses were conducted in Stata 15.1.[6] 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-.

3 | RESULTS

3.1 | Simulation studies

The results of the simulation studies (annual number of treated patients, $n = 5{,}000$), conducted using the rWTD with multiple random index dates, are presented in Table 1. Similar results were found for $n = 1{,}000$ and $n = 15{,}000$ and are hence not presented here (Appendix A, Table A1 and A2).

In Scenario 1, a correctly specified rWTD model, corresponding to a single Log-Normal IAD, was used to analyze the simulated data. With a single random index date, the RMSEs were 0.5 and 1.3 days for median prescription durations of 1.5 and 3 months, respectively. The precision was almost doubled by sampling five random index dates. In scenarios with median prescription durations of 3 months, the RMSE decreased from 1.3 days to 0.6 days when using 5 random index dates and further to 0.4 days with 20 random index dates. Whether sampling 1, 5, 10, or 20 random index dates for each patient, the relative bias was negligible (−0.04% to 0.01%) and the nominal coverage of 95% was almost retained (93.8%–95.4%) for all setups.

Scenario 2 introduces a form of length-bias due to the way treatment stopping is implemented (stopping is more likely to occur after a prescription with a long duration—for further details see Appendix A). Consequently, the relative bias (−0.43% to −0.20%) is slightly higher for all setups than in Scenario 1. As expected, the small bias is not affected by the number of random index dates used for each patient, and consequently the RMSE is the same as or only slightly higher than for Scenario 1 (0.8 and 0.6 days for median 3 months when sampling 5 and 20 random index dates, respectively). It shows the same tendency of RMSEs decreasing with the number of random index dates due to an increase in precision. As precision increases, the misspecification leads to lowered coverage probabilities (80.0% for median 3 months with 20 random index dates).

For Scenario 3 and 4 misspecification is more pronounced, since we use a model with a single Log-Normal IAD to fit data corresponding to a two- or three-component mixture of IADs. The varying degrees of misspecification lead to higher relative bias (−0.83% to 6.71%) and lower coverage probabilities (3.2%–95.2% for all setups using 1 random index date). The relative bias is unaffected by the number of random index dates and coverage probabilities decreases with the number of index dates (0.0%–95.1% using 20 random index dates). Again, the precision improves with the number of random index dates irrespective of the misspecification.

3.2 | Empirical study

The 80th percentiles for prescription durations of warfarin in Denmark, 2013–2014 based on stratified analyses using the rWTD with 1, 5, 10, and 20 random index dates for each patient showed improved precision with increasing number of index dates (Table 2). Using 1 random index date for each patient led to an estimate of 89.7 (87.7; 91.7) days for males corresponding to a width of the 95% confidence interval of 4.0 days. By increasing this to 5 random index dates for each patient, and hence increasing the total number of observations in the analysis fivefold, the precision improved with an estimate of 91.1 (90.0; 92.3) days, that is, the width decreased to 2.3 days. By increasing the number of random index dates for each individual further to 10 and 20, the precision improves slightly with the width decreasing to 2.0 and 1.9 days, respectively. For the subgroup of males aged 90 years or above, the width decreases from 35.4 to 16.2 days when using 5 random index dates instead of 1—and it further decreases to 11.1 days with 20 index dates. Similar tendencies for the precision where seen for the females and other age groups.

When a joint model for both sexes and all age groups was used to estimate prescription durations for groups defined by sex and age group, precision again increased with number of index dates (Table 3). We first explored how the three parameters of the rWTD (p, μ, and σ) depended on age and sex and found that the variation in prescription durations, σ, had no clear dependence on sex and age, while the proportion of patients continuing treatment, p, depended on age (for further details see Appendix B, Tables B1 and B2). Consequently, we present results from a model where σ did not depend on any covariates, p depended on age group, and μ depended on sex, age group and their interaction. Similar to the stratified analysis the width of the 95% confidence interval decreases for males aged 90 years or above, from 35.1 to 18.4 days, when using 5 random index dates instead of 1. It further decreases to 13.9 days with 20 index dates. Similar tendencies for the precision were seen for females and the other age groups.

4 | DISCUSSION

Using multiple random index dates for each patient to include more observations in the rWTD analysis increases the precision of the estimates as intended. When the model was correctly specified, precision doubled when using 5 random index dates instead of 1. In line with previous studies, the relative bias is negligible when there is no or very little misspecification.[4] It was a general finding that the relative bias of rWTD estimates of prescription duration was unaffected by the number of random index dates. Consequently, nominal coverage was retained in scenarios where the model was correctly specified, but in cases with misspecification, the coverage probabilities were lower when using more random index dates due to the increased precision. For the small strata in our empirical study, for example, males and females above 90 years, we have a low precision when using only 1 random index date and hence we see a more substantial gain in precision when using more random index dates here as compared to the larger strata. A stratified analysis based on sex and age group resembles a joint model adjusting all parameters for sex, age and their interaction. Results of the reduced model were similar with lowest precision using only 1 random index date, especially for the smaller groups of sex and age, and a precision that increased with number of random index dates.

TABLE 1 Simulations results for four different types of scenarios analyzed with the rWTD with a Log-Normal BRD

	IADs concerning stockpiling								1 random index date			5 random index dates			10 random index dates			20 random index dates		
Scenario	M1	VF1	M2	VF2	M3	VF3	MoS	True IAD 80th percentile (days)	RB (%)	RMSE (days)	CP (%)	RB (%)	RMSE (days)	CP (%)	RB (%)	RMSE (days)	CP (%)	RB (%)	RMSE (days)	CP (%)
1	1.5	1.5	-	-	-	-	-	54.3	−0.02	0.5	95.3	0.01	0.3	94.9	−0.00	0.2	95.3	−0.01	0.2	95.1
1	2	1.5	-	-	-	-	-	72.5	−0.01	0.8	94.6	−0.02	0.4	94.8	−0.01	0.3	93.8	−0.01	0.2	95.4
1	3	1.5	-	-	-	-	-	108.7	−0.04	1.3	95.1	−0.01	0.6	94.3	0.00	0.5	94.2	−0.00	0.4	94.4
2	1.5	1.5	-	-	-	-	-	54.3	−0.21	0.6	93.5	−0.21	0.3	92.8	−0.20	0.2	92.0	−0.21	0.2	88.0
2	2	1.5	-	-	-	-	-	72.5	−0.29	0.8	94.4	−0.28	0.4	91.2	−0.27	0.3	89.4	−0.26	0.3	86.6
2	3	1.5	-	-	-	-	-	108.7	−0.43	1.3	92.6	−0.41	0.8	88.4	−0.40	0.7	85.3	−0.41	0.6	80.0
3	1	2	1.5	2	-	-	2	44.1	−0.26	0.6	94.1	−0.25	0.3	93.5	−0.22	0.2	92.0	−0.22	0.2	91.1
3	1	1.5	1.5	1.5	-	-	2	39.1	1.00	0.6	86.2	1.00	0.4	53.6	1.00	0.4	27.5	1.00	0.4	8.9
3	1	1.5	1.5	2	-	-	2	38.6	2.75	1.2	43.8	2.77	1.1	0.3	2.78	1.1	0.0	2.77	1.1	0.0
3	1.5	2	2	2	-	-	2	64.9	−0.61	1.0	92.5	−0.62	0.6	86.2	−0.62	0.5	78.4	−0.61	0.5	72.4
3	1.5	1.5	2	1.5	-	-	2	57.6	0.02	0.6	95.2	0.08	0.3	94.8	0.08	0.2	94.6	0.09	0.2	95.1
3	1.5	1.5	2	2	-	-	2	57.4	1.41	1.1	81.9	1.42	0.9	35.0	1.43	0.9	15.2	1.43	0.9	4.6
3	1.5	2	3	2	-	-	2	71.0	0.33	1.2	95.2	0.33	0.7	93.2	0.35	0.6	92.3	0.31	0.5	92.1
3	1.5	1.5	3	1.5	-	-	2	61.7	6.15	3.9	3.2	6.15	3.8	0.0	6.12	3.8	0.0	6.12	3.8	0.0
3	1.5	1.5	3	2	-	-	2	60.8	6.71	4.2	4.8	6.67	4.1	0.0	6.70	4.1	0.0	6.68	4.1	0.0
4	1.5	1.5	1	1.5	2	1.5	2	55.0	−0.20	0.6	94.4	−0.16	0.3	93.3	−0.17	0.2	92.6	−0.17	0.2	91.0
4	1.5	1.5	1	1.5	2	1.5	3	55.3	−0.49	0.7	91.8	−0.47	0.4	86.0	−0.47	0.3	80.0	−0.47	0.3	71.1
4	3	1.5	2	1.5	4	1.5	2	106.3	−0.44	1.4	92.5	−0.40	0.8	89.2	−0.41	0.7	85.4	−0.41	0.6	80.8
4	3	1.5	2	1.5	4	1.5	3	108.2	−0.81	1.6	89.0	−0.83	1.1	73.6	−0.81	1.0	64.1	−0.83	1.0	50.0

Note: No stockpiling (Scenario 1 and 2) and stockpiling at the end of the year due to either larger redemptions (Scenario 3) or more frequent redemptions (Scenario 4). Stockpiling can occur in the last 2 or 3 months of the year. For all four scenarios on average 25% are stopping treatment within a one-year period. For Scenario 1 there is a constant probability of ending treatment and for Scenario 2, 3, and 4 the duration of the treatment episodes are Log-Normally distributed with a mean of 3 years and a standard deviation of 2 years. Prescriptions redeemed in periods without and with stockpiling follows IAD_1 and IAD_2, respectively. For Scenario 4 the last prescription after stockpiling follows IAD_3. All IADs are Log-Normal and are specified by their median and variation factor. A variation factor of 1.5 means that the 97.5% percentile of the density is 1.5 times the median. The simulated data for the different scenarios are all analyzed using the rWTD with 1, 5, 10, and 20 random index dates for each patient, respectively. For each scenario 2500 datasets were simulated with an annual number of treated patients of 5000. *M* is the median in months. *VF* is the variation factor. *MoS* is the number of months of stockpiling. *RB* is the relative bias and is computed as the average relative difference between the estimated 80th percentile and the true 80th percentile. *RMSE* is the root mean square error of the estimates around the true value of the 80th percentile. *CP* is the coverage probability defined by the percentage of nominal 95% confidence intervals, obtained by the normal approximation, containing the true 80th percentile.

TABLE 2 Estimated 80th percentiles for prescription durations of warfarin in 2013–2014 based on a stratified analysis using the rWTD with 1, 5, 10, and 20 random index dates for each patient, respectively

			1 random index date			5 random index dates			10 random index dates			20 random index dates		
Sex	Age (years)	n*	n_{pres}**	Estimated 80th percentile in days (95% CI***)	Width CI (days)	n_{pres}	Estimated 80th percentile in days (95% CI)	Width CI (days)	n_{pres}	Estimated 80th percentile in days (95% CI)	Width CI (days)	n_{pres}	Estimated 80th percentile in days (95% CI)	Width CI (days)
Female	0–49	843	635	75.8 (67.9; 83.8)	15.9	3165	74.4 (69.0; 79.8)	10.8	6341	76.8 (72.2; 81.5)	9.3	12 662	77.2 (73.1; 81.4)	8.3
Female	50–59	641	520	79.9 (70.3; 89.6)	19.3	2559	79.4 (74.2; 84.6)	10.4	5114	80.0 (75.3; 84.6)	9.3	10 205	79.9 (75.7; 84.2)	8.5
Female	60–69	1617	1315	83.6 (77.9; 89.3)	11.4	6619	84.9 (81.6; 88.2)	6.6	13 252	86.4 (83.4; 89.3)	5.9	26 562	86.1 (83.4; 88.7)	5.3
Female	70–79	3228	2763	96.5 (92.1; 101.0)	9.0	13 880	96.2 (93.7; 98.8)	5.1	27 720	96.7 (94.5; 98.9)	4.4	55 428	96.4 (94.4; 98.3)	4.0
Female	80–89	3384	2887	111.3 (105.8; 116.8)	11.0	14 429	106.9 (104.0; 109.9)	5.9	28 911	107.0 (104.5; 109.5)	5.0	57 717	107.0 (104.7; 109.2)	4.5
Female	90+	854	693	124.2 (110.1; 138.3)	28.1	3463	124.0 (116.5; 131.5)	15.0	6934	121.6 (115.2; 128.1)	12.8	13 908	118.0 (112.5; 123.5)	11.0
Female	All	10 567	8785	98.6 (95.9; 101.3)	5.4	44 116	97.9 (96.4; 99.5)	3.1	88 204	97.5 (96.1; 98.8)	2.7	176 388	97.4 (96.2; 98.7)	2.5
Male	0–49	902	686	74.7 (66.2; 83.2)	17.0	3485	74.8 (70.2; 79.5)	9.3	6960	75.0 (70.9; 79.1)	8.2	13 932	74.5 (70.8; 78.1)	7.3
Male	50–59	1361	1093	79.6 (73.3; 85.9)	12.6	5488	79.4 (75.5; 83.3)	7.8	10 950	79.4 (76.1; 82.7)	6.6	21 862	79.5 (76.5; 82.4)	5.9
Male	60–69	3474	2963	88.7 (84.5; 92.8)	8.3	14 793	87.4 (85.0; 89.7)	4.6	29 504	87.8 (85.8; 89.8)	4.0	59 097	87.4 (85.6; 89.2)	3.6
Male	70–79	5308	4600	93.6 (90.3; 96.8)	6.5	23 058	94.1 (92.2; 96.0)	3.8	46 085	93.8 (92.2; 95.5)	3.2	92 189	94.3 (92.7; 95.8)	3.0
Male	80–89	3572	3104	96.9 (92.3; 101.4)	9.1	15 477	97.7 (95.2; 100.3)	5.0	30 907	98.3 (96.1; 100.6)	4.4	61 774	98.5 (96.5; 100.5)	4.0
Male	90+	517	406	99.9 (82.2; 117.6)	35.4	2084	104.5 (96.4; 112.6)	16.2	4142	103.0 (96.4; 109.5)	13.1	8246	102.3 (96.7; 107.9)	11.1
Male	All	15 134	12 881	89.7 (87.7; 91.7)	4.0	64 418	91.1 (90.0; 92.3)	2.3	128 718	91.7 (90.6; 92.7)	2.0	257 328	91.7 (90.7; 92.6)	1.9
All		25 701	21 688	93.4 (91.8; 95.0)	3.2	108 344	93.8 (92.9; 94.8)	1.9	216 858	93.6 (92.8; 94.4)	1.6	433 811	93.6 (92.9; 94.4)	1.5

Note: A Log-Normal BRD is used for fitting the prevalent component. n*: number of patients redeeming at least one prescription of warfarin in 2013–2014. n_{pres}**: number of prescriptions sampled in the analysis (with replacement). CI***: For ease of comparison all confidence intervals are calculated on the time scale and not the logarithmic timescale although, their asymptotic properties are suboptimal in the small subgroups when using only one index date.

TABLE 3 Estimated 80th percentiles for prescription durations of warfarin in 2013–2014 based on the rWTD with 1, 5, 10, and 20 random index dates for each patient, respectively

		1 random index date (n_{pres}^{*} = 21 688)		5 random index dates (n_{pres} = 108 344)		10 random index dates (n_{pres} = 216 858)		20 random index dates (n_{pres} = 433 811)	
Sex	Age (years)	Estimated 80th percentile in days (95% CI**)	Width CI (days)	Estimated 80th percentile in days (95% CI)	Width CI (days)	Estimated 80th percentile in days (95% CI)	Width CI (days)	Estimated 80th percentile in days (95% CI)	Width CI (days)
Female	0–49	81.1 (72.9; 89.3)	16.4	76.5 (71.3; 81.6)	10.2	77.0 (72.6; 81.3)	8.7	77.4 (73.2; 81.5)	8.3
Female	50–59	79.0 (70.0; 88.0)	18.1	77.4 (72.1; 82.7)	10.6	78.6 (74.0; 83.2)	9.2	79.5 (75.2; 83.8)	8.6
Female	60–69	88.2 (82.3; 94.1)	11.8	88.4 (84.9; 91.9)	7.0	86.4 (83.4; 89.4)	6.0	86.4 (83.6; 89.2)	5.6
Female	70–79	97.2 (93.0; 101.5)	8.5	97.4 (94.8; 100.0)	5.1	97.1 (94.8; 99.4)	4.6	97.0 (94.8; 99.1)	4.3
Female	80–89	103.6 (99.0; 108.2)	9.2	106.2 (103.5; 108.9)	5.5	106.7 (104.2; 109.1)	4.9	107.0 (104.8; 109.3)	4.6
Female	90+	109.5 (97.3; 121.6)	24.3	115.8 (108.8; 122.9)	14.1	115.2 (108.9; 121.4)	12.5	114.0 (108.5; 119.6)	11.1
Male	0–49	71.3 (63.9; 78.7)	14.9	74.3 (70.0; 78.7)	8.8	73.9 (70.0; 77.7)	7.7	74.6 (70.9; 78.3)	7.4
Male	50–59	81.7 (75.3; 88.2)	12.8	80.1 (76.4; 83.9)	7.5	80.8 (77.3; 84.2)	6.9	80.0 (76.9; 83.1)	6.2
Male	60–69	86.6 (82.9; 90.3)	7.4	87.5 (85.2; 89.7)	4.6	87.2 (85.2; 89.1)	4.0	86.6 (84.8; 88.4)	3.6
Male	70–79	95.1 (92.1; 98.2)	6.1	93.9 (92.0; 95.9)	3.8	94.2 (92.5; 95.9)	3.4	94.1 (92.5; 95.6)	3.1
Male	80–89	97.3 (93.2; 101.4)	8.2	98.0 (95.6; 100.5)	4.9	97.3 (95.1; 99.4)	4.4	98.1 (96.0; 100.1)	4.1
Male	90+	106.4 (88.9; 124.0)	35.1	110.4 (101.2; 119.6)	18.4	107.6 (100.2; 114.9)	14.7	107.6 (100.6; 114.6)	13.9

Note: A Log-Normal BRD with parameters μ and σ is used for fitting the prevalent component. In the rWTD p depends on age group, μ on sex, age group and their interaction, and σ does not depend on any covariates. p: fraction of prevalent users among the observed users in the observation window. n_{pres}*: number of prescriptions sampled in the analysis (with replacement). CI**: For ease of comparison all confidence intervals are calculated on the time scale and not the logarithmic timescale although, their asymptotic properties are suboptimal in the small subgroups when using only one index date.

The primary advantage of sampling multiple random index dates for each patient is that we utilize the information in the data more efficiently than when sampling only one random index date. The approach is straightforward since we for each random index date consider the time since the last preceding prescription redemption and use them as before in the rWTD analysis to obtain estimates of the 80th percentile of the prescription duration. Hence, by sampling multiple random index dates we have a larger dataset to perform the rWTD analysis on, which improves the precision of the estimates. However, we now have to use robust variance estimation to allow for the dependence of multiple observations from the same individual.

This further development of the rWTD method shares most limitations with the previous versions using a single index date, such as ignoring censoring, being applicable for drugs with a more or less chronic usage pattern and being sensitive to misspecification due to seasonality in rates of prescription redemptions or treatment stopping.[1–4] The sensitivity of the model to misspecification, combined with the improved precision, leads to lower coverage probabilities when the number of random index dates is increased. It is therefore crucial to investigate the model fit in diagnostic plots to detect misspecification. However, even though bias may be present, it is reassuring that even for the most extreme case with stockpiling due to larger redemptions, the estimate was only 4.1 days off from the true duration of 60.8 days, which may in many applications be within the acceptable margin of error.

When prescription durations are not recorded directly in the pharmacoepidemiological databases they have to be estimated, either from the available data or by imposing external knowledge. This is the case for many countries, for example, Denmark, but for some countries, for example, in North America, databases contain measures of days supply.[5,7] However, although intended days supply are registered in the databases this is, not necessarily, consistent with their actual consumption and hence it can still be relevant to have a way of estimating prescription durations. The WTD approach is intended to estimate prescription durations consistent with actual consumption. Prescription durations are directly associated with treatment durations which in turn are used for defining exposure status. Hence the choice of prescription duration can drastically affect the results of a pharmacoepidemiologic study. A common approach for estimating prescription durations has been to use decision rules, such as assuming a consumption of one defined daily dose or one tablet per day, or other more flexible decision rules.[8–11] These approaches are based on decision algorithms and it is not straightforward to validate and refine these methods since they are not based on an explicit model, in contrast to the WTD-based approaches. The main challenge to using the WTD has until now been its poor efficiency, as it only utilizes one prescription for each patient. Using multiple random index dates improves precision, which is a virtue, although it does not affect bias and hence causes coverage probabilities to drop when the model is misspecified. In general using the WTD with random index dates is preferable since this method performs equally well or even better than the WTD with a fixed index date.[4] However, it also depends on whether we are interested in a marginal estimate for the entire period or the point prevalence at a given time. For the latter we need to use the WTD with a fixed index date, assuming there is no misspecification.

Using the WTD with more random index dates yields a considerable gain in precision, which becomes even more evident when we have a small sample size or when we wish to account for patient characteristics in stratified or adjusted analyses. However, this gain in precision comes at a computational cost. For the scenarios without seasonal variation (Scenario 1 and 2) we found that the RMSE was almost reduced by half when sampling 5 random index dates instead of 1. Using $j \geq i$ random index dates leads to an approximate $\sqrt{j/i}$ times decrease in standard error as compared to using i random index dates. Sampling 5 random index dates instead of 1 thus gives us approximately 2.2 times narrower confidence intervals and 20 random index dates 4.5 times narrower. Even though precision improves with the number of random index dates, the most noticeable improvement is obtained with the first additional index dates, that is, increasing from 1 to 5. Since the number of observed prescription redemptions is limited for each patient, there is likely to be an upper limit for the number of index dates that can be meaningfully included. However, even with 20 index dates we found a small improvement in precision, although this exceeded the average annual number of individual redemptions.

In conclusion, we suggest using the rWTD with multiple random index dates to obtain estimates of the marginal prescription duration over a sampling period of interest with a higher precision. Further work is needed on how to detect and account for misspecification, for example by inclusion of covariates in the model and we expect that having more data available for the analyses will give us a higher flexibility in trying to model various kinds of misspecification.

ACKNOWLEDGMENTS

We would like to thank Professor Anton Pottegård (Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, Odense, Denmark) for helpful input and comments to draft versions of the paper.

CONFLICT OF INTEREST

The authors declare no personal conflicts of interest pertaining to this work. Department of Clinical Epidemiology is, however, involved in studies with funding from various companies as research grants to (and administered by) Aarhus University.

ETHICS STATEMENT

According to Danish law observational studies based on routinely collected health care data do not require ethical approval.

ORCID

Katrine Bødkergaard https://orcid.org/0000-0001-9128-0383
Jesper Hallas https://orcid.org/0000-0002-8097-8708
Eva Skovlund https://orcid.org/0000-0002-2997-6141
Henrik Støvring https://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. Estimating medication stopping fraction and real-time prevalence of drug use in pharmacoepidemiologic databases. An application of the reverse waiting time distribution. *Pharmacoepidemiol Drug Saf*. 2017;26(8):909-916.
3. 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;26(8):900-908.
4. Bødkergaard K, Selmer RM, Hallas J, et al. Using the waiting time distribution with random index dates to estimate prescription durations in the presence of seasonal stockpiling. *Pharmacoepidemiol Drug Saf*. 2020;29(9):1072-1078.
5. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data resource profile: the Danish national prescription registry. *Int J Epidemiol*. 2017;46(3):798-798f.
6. StataCorp. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP; 2015.
7. The Centers for Medicare & Medicaid Services. Medicare Fee-For Service Provider Utilization & Payment Data Part D Prescriber Public Use File: A Methodological Overview. 2020. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Downloads/Prescriber_Methods.pdf. Accessed May 20, 2021.
8. Mantel-Teeuwisse AK, Klungel OH, Verschuren WMM, Porsius 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.
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(4):422-427.
10. 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):1-13.
11. Williams R, Brown B, Peek N, Buchan I. Making medication data meaningful: illustrated with hypertension. *Stud Health Technol Inform*. 2017;228:247-251.

SUPPORTING INFORMATION

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

How to cite this article: Bødkergaard K, Selmer RM, Hallas J, Kjerpeseth LJ, Skovlund E, Støvring H. Using multiple random index dates with the reverse waiting time distribution improves precision of estimated prescription durations. *Pharmacoepidemiol Drug Saf*. 2021;30(12):1727-1734. doi: 10.1002/pds.5340