



Impact of "time zero" of Follow-Up Settings in a Comparative Effectiveness Study Using Real-World Data with a Non-user Comparator: Comparison of Six Different Settings

Ryozo Wakabayashi^{1,2} · Takahiro Hirano^{1,2} · Thomas Laurent^{1,2} · Yoshiki Kuwatsuru³ · Ryohei Kuwatsuru^{1,3}

Accepted: 6 November 2022 / Published online: 28 November 2022
© The Author(s) 2022

Abstract

Background Time-related bias can lead to misleading conclusions. Properly setting the "time zero" of follow-up is crucial for avoiding these biases. However, the time-zero setting is challenging when comparing users and non-users of a study drug because the latter do not have a time point for starting treatment.

Objective This methodological study aimed to illustrate the impact of different time-zero settings on effect estimates in a comparative effectiveness study using real-world data with a non-user comparator.

Methods Data for type 2 diabetes patients were extracted from an administrative claims database, and the onset of diabetic retinopathy (study outcome) was compared between users (treatment group) and non-users (non-use group) of lipid-lowering agents. We applied six time-zero settings to the same dataset. The adjusted hazard ratio (HR) for the outcome was estimated using a Cox regression model in each time-zero setting, and the obtained results were compared among the settings.

Results Of the six settings, three (study entry date [SED] vs SED [naïve approach], treatment initiation [TI] vs SED, TI vs Matched [random order]) showed that the treatment had a reduced risk of the outcome (HR [95% CI]: 0.65 [0.61–0.69], 0.92 [0.86–0.97], and 0.76 [0.71–0.82], respectively), one (TI vs Random) had an increased risk (HR [95% CI]: 1.52 [1.40–1.64]), and two (SED vs SED [cloning method], and TI vs Matched [systematic order]) had neither increased nor decreased risk (HR [95% CI]: 0.95 [0.93–1.13], and 0.99 [0.93–1.07], respectively).

Conclusions This study demonstrates that different time-zero settings can lead to different conclusions, even if the same dataset is analyzed for the same research question, probably because improper settings can introduce bias. To minimize such biases, researchers should carefully define time zero, particularly when designing a non-user comparator study using real-world data.

Key Points

This methodological study compared the effect estimates among different time-zero settings in a comparative effectiveness study with a non-user comparator using actual real-world data.

The treatment effects, based on the adjusted hazard ratio for the outcome, varied among the six time-zero setting methods, even though the same dataset was used for the same research question.

An incorrectly set time zero can result in a biased result and mislead the conclusion; therefore, researchers should carefully select the time zero to minimize those biases when designing a study.

✉ Ryozo Wakabayashi
ryo_wakabayashi@jp-css.com

¹ Real-World Evidence and Data Assessment (READS), Graduate School of Medicine, Juntendo University, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan

² Clinical Study Support, Inc., Nagoya, Japan

³ Department of Radiology, School of Medicine, Juntendo University, Tokyo, Japan

1 Introduction

An increasing body of pharmacoepidemiologic research has been conducted using real-world data (RWD). Real-world data studies can generate evidence on the effectiveness or safety of a treatment in clinical settings, which complements the evidence obtained from randomized controlled trials [1]. Moreover, RWD can be utilized in research throughout the life cycle of drug development, including regulatory decision making [2, 3]. Real-world data studies have research advantages, such as representativeness, large data size, longitudinal data, and low cost [4]. However, using RWD also has serious challenges because of potential biases, such as misclassification of exposure and outcomes, confounding, and selection bias, albeit some inherent in observational studies [4].

Time-related biases, such as immortal-time bias and time-lag bias, are major issues in comparative effectiveness studies using RWD [5, 6]. Immortal time is a period during which the outcome cannot occur because of the definition of the cohorts [7]. For example, patients in a treatment group inevitably remain event-free until they receive the study treatment, and inappropriate handling of this period can introduce bias in the result (immortal-time bias) [8, 9]. Time-lag bias can arise when comparing groups at different stages of the disease, for instance, when the time from diagnosis to the start of follow-up differs between groups [8]. These time-related biases can significantly impact the estimates of treatment effects, and failure to control for them may lead to misleading conclusions [5, 10].

Time-related biases are typically introduced when baseline or follow-up-related time points differ between the study groups. Therefore, aligning the three time points—meeting eligibility, initiating treatment, and starting follow-up—is crucial to reduce these biases in a comparative effectiveness study [11, 12]. One solution is to use a new-user, active-comparator design, that is, comparing new users of treatment of interest with new users of alternative therapy (ideally with the same indication) [1, 13]. Using this comparator instead of prevalent users or patients without active comparators (non-users) can reduce selection bias and increase between-group comparability [13, 14]. Moreover, in the new-user design, the three time points described above (eligibility, treatment start, and follow-up start) can be aligned by setting the starting point of follow-up (“time zero”) at treatment initiation (TI) in both groups, thereby reducing the time-related biases [11].

Researchers may want to compare against non-users for their study purposes, or because of the absence of a proper active comparator. Indeed, of 89 studies on drug effectiveness or safety using RWD, 34% used a non-user comparator [12]. However, comparison with non-users poses a significant challenge in setting the time zero because non-users do

not have a time point for starting treatment [12]. Some solutions related to time-zero problems when compared to non-users were proposed [11] and previous studies selected time zero for the non-use group in various ways: for example, at cohort entry, which is the same for the treatment group [15]; or at a different time point compared to the treatment group, including the first day of the first diagnosed month [16], a randomly selected visit date [17], and the same date as their matched patient [18]. As such, various options are available for selecting time zero, but the proper selection is crucial to avoid time-related bias [11, 12]. There are studies using simulated data [19] or RWD [20], which examined various approaches for selecting time zero, although they focused on time-zero setting of external control cohort with multiple line initiation as potential time zeros in oncology.

However, the degree to which the study results would be affected by the different time-zero settings using RWD, especially in a non-cancer study with a non-user comparator, has not been thoroughly examined. Therefore, this study aimed to illustrate the impact of “time-zero setting,” the setting of the starting point of follow-up, on the effect estimates in a cohort study using RWD to compare the treatment effect between the treatment and non-use groups.

2 Methods

2.1 Study Overview

This was a methodological study to illustrate the impact of different settings of time zero of follow-up in a pharmacoepidemiological, non-user comparator study using actual RWD. We conceived several methods for setting time zero and compared the effect estimates obtained in each setting. As an example of an RWD study on treatment effectiveness, we used a previous cohort study using Japanese RWD that compared the risk of diabetic retinopathy (DR) between users and non-users of lipid-lowering medications in patients with type 2 diabetes [21]. The present study was approved by the ethics committee of Juntendo University, Tokyo, Japan (approval number: E21-0041).

2.2 Data Source

This study used an administrative claims database compiled by JMDC Inc., one of Japan's most frequently used RWD sources [22]. As of September 2021, the number of patients in the database is approximately 13 million. The database contains health checkup results and claims (for hospitalization, outpatient treatment, drug preparation, and dental treatment) for insured employees and their dependents collected from health insurance societies [23]. The database includes

patient attributes (age and sex), medical facilities, diagnoses, procedures, medications, and medical costs. In this database, data are available from 2005, and anonymized individual IDs enable the tracing of patients between institutions so long as they are insured by the same insurance society.

2.3 Study Cohort

The study cohort comprised patients with type 2 diabetes, initiated on glucose-lowering medications between January 1, 2005, and March 31, 2017 (study period). The study entry date (SED) was defined as the date of the first prescription of a glucose-lowering agent (Anatomical Therapeutic Chemical [ATC] classification system code, A10) and with the diagnostic record of type 2 diabetes (International Classification of Diseases [ICD]-10 codes, E11 or E14) in the same month. To check for pre-entry patient characteristics to confirm eligibility, we required patients to have data for at least 180 days before SED.

Patients were excluded if they had a prescription record of lipid-lowering agents (ATC code, C10) or a diagnostic record of type 1 diabetes (ICD-10 code, E10) during the 180 days before SED or if they were aged < 18 years at SED. A history of DR before SED was handled by setting a washout period for the outcomes (described in Sect. 2.5).

2.4 Variables

The selection and definition of variables were based on a study by Kawasaki et al [21].

2.4.1 Outcome

The outcome was identified using a diagnostic record of DR (ICD-10 codes: E103, E113, E143, H360, H431, H356, H450, or H431). The event date was defined as the date of the first diagnostic record of DR in the patient data.

2.4.2 Exposure

Patients were classified into the treatment group if they had a prescription record of lipid-lowering agents (ATC code C10) during the period from SED to the end of follow-up (defined in Sect. 2.5) or otherwise into the non-use group.

2.4.3 Covariates

The covariates were age, sex, and duration of type 2 diabetes at SED, in addition to the following variables assessed over 180 days before SED: weighted Charlson Comorbidity Index [24], antihypertensive medications (ATC codes, C02, C03, or C07–09), and anticoagulant agents (ATC codes, B01AA-AF or B01AX).

2.5 Follow-Up and Wash-Out Periods

We implemented several methods for setting “time zero,” the starting point of the follow-up. Each method is presented in Sect. 2.6. Regardless of the method used, the follow-up period began at time zero and ended when censored on the incidence of an outcome event, exit from the database, or the end of the study period, whichever occurred first. For simplicity, treatment discontinuation was not considered for the treatment group. In other words, the “as-started effect” (the effect of the initial treatment choice) was evaluated without considering the treatment duration [1].

We set a washout period for the outcomes, during which an outcome event should not occur, as “before time zero.” As the time point of time zero varies depending on the setting method, the washout period also changes according to the settings. For example, when time zero was set at SED, patients with an outcome event at any time before SED were excluded. However, when time zero was set at TI, patients with an outcome event after SED but before TI were also excluded.

2.6 “Time Zero” Settings

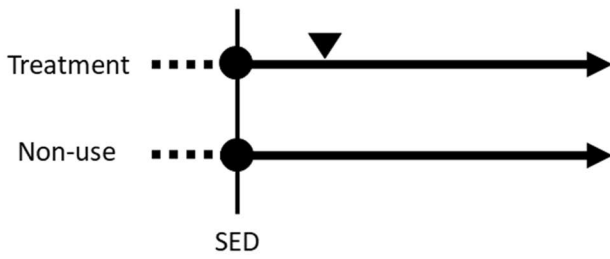
We conceived four different settings of time zero that were previously proposed or used (SED vs SED, TI vs SED, TI vs Random, and TI vs Matched), of which SED versus SED and TI versus Matched had two approaches; thus, a total of six time-zero settings were implemented in this study. The study timeframe and definition of time zero in the respective settings are displayed in Fig. 1.

2.6.1 SED versus SED

Time zero was set at SED for both the treatment and non-use groups (Fig. 1a). Two approaches were used to implement this setting. The first used the “naïve approach”, in which patients who had an outcome event before TI were treated as non-users [15].

The second approach used the “cloning method”, which was proposed to address biases caused by misalignment of follow-up and treatment starting points [11, 14]. Step-by-step instructions for this method have been published elsewhere [25]. In brief, two clones (copies) were created for each patient; one was allocated to the treatment arm and the other to the non-use arm. Either clone was then censored when it became incompatible with the definition of the allocated arm (e.g., a prescription of lipid-lowering agents in the non-use arm). This artificial censoring introduces a selection bias. Therefore, inverse-probability-of-censoring weighting [26] was applied to address this informative censoring. Weights were calculated using a Cox regression model, including all covariates (listed in the Sect. 2.4.3).

a. SED vs SED



Time zero

- Treatment: at SED
- Non-use: at SED

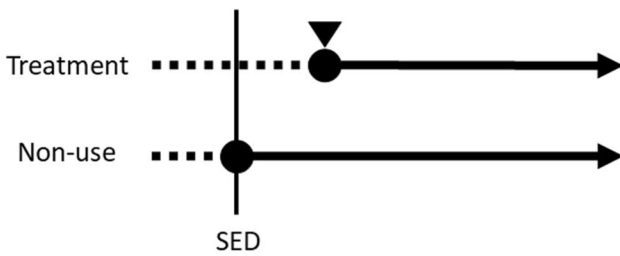
SED vs SED (naïve approach)

- Patients who had an outcome event before treatment initiation were treated as non-users.

SED vs SED (cloning method)

- Two clones were created for each patient, and each was allocated to the respective group.

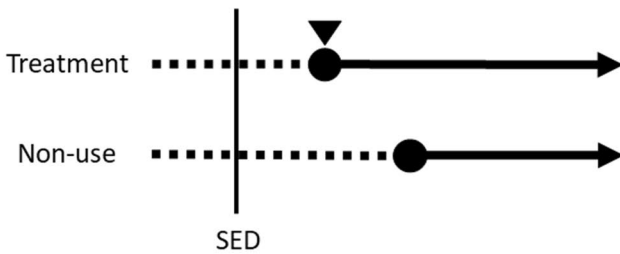
b. TI vs SED



Time zero

- Treatment: at treatment initiation
- Non-use: at SED

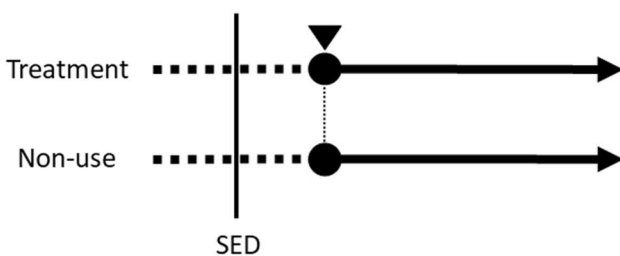
c. TI vs Random



Time zero

- Treatment: at treatment initiation
- Non-use: at random hospital visit date after SED

d. TI vs Matched



Time zero

- Treatment: at treatment initiation
- Non-use: at the date corresponding to treatment initiation of the matched patient

TI vs Matched (random order)

- Patients were matched in random order.

TI vs Matched (systematic order)

- Patients were matched systematically based on the duration from SED to treatment initiation

● = Time zero ▼ = Treatment initiation (1st prescription date of lipid-lowering agents)
 ————— = Follow-up period = Washout period for the outcomes

In contrast to the original method, in this study, we did not consider a grace period, and consequently, among the patients in the treatment arms, patients in both groups were

not artificially censored since they were always compatible with the arm. For the non-use arm, patients in the treatment groups were artificially censored at treatment initiation,

Fig. 1 Study time frame for each time-zero setting method. The follow-up period started at time zero and ended when censored on the incidence of an outcome event, exit from the database, or the end of the study period, whichever occurred first. The time before time zero was the washout period for the outcomes; patients with an outcome event during the washout period were excluded from the analysis. *SED* study entry date (the date of the first prescription of a glucose-lowering agent plus the diagnostic record of type 2 diabetes in the same month), *TI* treatment initiation (first prescription date of lipid-lowering agents)

i.e., the time when they were not compatible with the arm, although those in the non-use groups were not. Hence, the events in non-use groups were considered in both arms. Since we suspected that this may dilute the treatment effect, we performed an analysis in the per-protocol setting using different grace periods as mentioned in the supplementary materials (Supplementary materials).

2.6.2 TI versus SED

Time zero was set in the treatment group at TI, the first prescription date of lipid-lowering agents [16]. In contrast, it was set at SED for the non-use group (Fig. 1b). In this setting, patients with an outcome event before TI were excluded from the treatment group because the outcome occurred within the washout period.

2.6.3 TI Versus Random

Time zero for the treatment group was set at TI, as in TI versus SED. In contrast, for the non-use group, the date of time zero was randomly selected from all hospital visit dates after SED (Fig. 1c) [11, 17].

2.6.4 TI versus Matched

Time zero for the treatment group was set at TI. In contrast, time zero for the non-use group was defined as the corresponding date of time zero (in terms of the length from SED to TI) of their matched treatment patients. Therefore, the duration from SED to time zero was equal between the matched pairs (Fig. 1d). To implement this setting, treatment patients were matched with non-users in a 1:1 ratio using a propensity score for lipid-lowering treatment calculated using a logistic regression model including all covariates (listed in the Sect. 2.4.3).

We used two approaches for matching, differing in the order of patient matching. The first approach matched patients in random order (TI vs Matched [random order]). Patients in the treatment group were listed randomly, and matching was performed sequentially from the top of the list [27]. For each treatment patient, a non-user with the nearest propensity score was identified as a potential match. The time zero of this non-user was set to correspond to the

duration from SED to the time zero of the matched treatment patient. A non-user was discarded if this patient had been censored by this date, and matching was repeated until a suitable non-user was identified. However, this non-user could still be a potential match for another treatment patient.

The second approach matched patients in a systematic order (TI vs Matched [systematic order]). Patients in the treatment group were listed in ascending order on the length from SED to TI and matching was performed sequentially from the top of the list (i.e., from treatment patient who had shorter length from SED to TI) [28]. As for the TI versus Matched (random order), for each treatment patient, a non-user with the nearest propensity score was identified as a potential match and time zero was set to correspond to the duration from SED to the time zero of the matched treatment patient. A patient was discarded when censored by this date. In the systematic order matching, a non-user, who was once discarded due to censoring before the time zero of the potential match, was also discarded from subsequent matching.

2.7 Statistical Analyses

In each time-zero setting, the following effect measurements were calculated: (1) cumulative incidence at 1 year from time zero (2) incidence rate of DR; (3) crude incidence rate ratio of DR; and (4) hazard ratio (HR) and its 95% confidence interval (CI), estimated using a Cox regression model adjusted for covariates (listed in the Sect. 2.4.3). However, for SED versus SED (cloning method), only HR was estimated because the calculation of the incidence rate and crude incidence rate ratio was not applicable due to the use of an artificial population in the cloning method. In addition, only the group (treatment or non-use) was included as an explanatory variable in the Cox regression model for the outcome because covariate adjustment was already performed in inverse-probability-of-censoring weighting.

As a sensitivity analysis, the effect measurements were also calculated after excluding patients with an outcome event within 90 days after time zero in each method. Analyses were performed using SAS release 9.4 (SAS Institute, Cary, NY, USA), and R version 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria). The latter was used for analysis of the SED versus SED (cloning method) only.

3 Results

3.1 Baseline Characteristics of Patients

Of 12,706,548 individuals with ICD-10 codes of E11–14 available in the JMDC database at the time of the study, we obtained data for 42,308 patients who met the eligibility

criteria. The flow of patient disposition is provided in the Online Resource Fig. S1.

These eligible patients were classified into treatment or non-use groups, and each time-zero setting was applied. In each method, patients with an outcome event during the washout period, which differed depending on the time-zero settings, were excluded. As a result, 10,929 patients comprised the treatment group in all time-zero settings except for SED versus SED (cloning method); and 24,767, 23,935, 21,860, and 10,929 patients formed the non-use groups in SED versus SED (naïve approach), TI versus SED, TI versus Random, and TI versus Matched (random order, systematic order), respectively. The baseline characteristics of the patients in each group are summarized in Table 1.

The treatment group in TI versus SED, TI versus Random, and TI versus Matched (time zero at TI) had a mean (standard deviation [SD]) duration from SED to time zero of 154.4 (345.5) days. In contrast, the non-use group had 0 days in TI versus SED (time zero at SED) and 306.6 (435.0) days in TI versus Random (time zero at randomly sampled hospital visit date).

3.2 Relative Risk of Diabetic Retinopathy

Table 2 and Fig. 2 summarize the effect measurements in each time-zero setting.

In SED versus SED (naïve approach), the cumulative incidence at 1 year follow-up and its 95% CI of treatment group was 0.070 (0.065–0.075) and that of the non-use group was 0.124 (0.119–0.128). In SED versus SED (cloning method), that of the treatment group was 0.101 (0.097–0.104) and that of the non-use group was 0.100 (0.096–0.103). In TI versus SED, that of the treatment group was 0.087 (0.081–0.092), which was same as in TI versus Random, TI versus Matched (random order) and TI versus Matched (systematic order), and that of the non-use group was 0.098 (0.094–0.103). Those of the non-use group were 0.054 (0.050–0.058) in TI versus Random, 0.122 (0.115–0.129) in TI versus Matched (random order), and 0.090 (0.084–0.096) in TI versus Matched (systematic order).

In SED versus SED (naïve approach), the incidence rate per 1000 person-year and its 95% CI of treatment group was 57.8 (55.0–60.6) and that of the non-use group was 94.9 (91.9–97.8). In TI versus SED, that of the treatment group was 69.1 (65.8–72.5), which was same as in TI versus Random, TI versus Matched (random order) and TI versus Matched (systematic order), and that of the non-use group was 76.5 (73.8–79.1). Those of the non-use group were 51.6 (48.6–54.6) in TI versus Random, 100.8 (95.9–105.7) in TI versus Matched (random order), and 72.2 (68.5–75.9) in TI versus Matched (systematic order).

Among all methods, both the crude rate ratio and adjusted HR showed the same trends. However, the results of the

treatment effects were inconsistent among the six methods. Lipid-lowering treatment was associated with a reduced risk of DR in SED versus SED (naïve approach) (HR, 0.65; 95% CI 0.61–0.69), TI versus SED (RR, 0.61; 95% CI 0.58–0.65, adjusted HR, 0.92; 95% CI 0.86–0.97), and TI versus Matched (random order) (RR, 0.69; 95% CI 0.64–0.73, adjusted HR, 0.76; 95% CI 0.71–0.82). In contrast, TI versus Random showed that treatment was associated with an increased risk (RR, 1.34; 95% CI 1.24–1.44, adjusted HR, 1.52; 95% CI 1.40–1.64). Contrary to these results, treatment had neither increased nor decreased risk in SED versus SED (cloning method) (adjusted HR, 0.95; 95% CI 0.93–1.13) and TI versus Matched (systematic order) (RR, 0.96; 95% CI 0.89–1.03, adjusted HR, 0.99; 95% CI 0.93–1.07).

In this population, outcome events occurred more frequently in the early days after SED (Online Resource Fig. S2). Therefore, we calculated the effect estimates after excluding patients with an outcome within 90 days after time zero as a sensitivity analysis. Although the effect estimates approached null in every time-zero setting, each method showed the same trend as above, except that TI versus SED no longer showed any treatment effect (Online Resource Table S1).

For SED versus SED (cloning method), we performed an analysis in the per-protocol setting using different grace periods (either 3 or 6 months) and did not observe any differences in a sensitivity analysis (supplementary materials).

4 Discussion

In a comparative effectiveness study, multiple choices are available for when and how to set the time zero of follow-up [11, 19, 20], and the time-zero setting is particularly challenging in a study with a non-user comparator arm [12]. In this study, we implemented six methods to set time zero in a comparative RWD study with a non-user comparator and compared the effect estimates obtained from these settings. The six methods estimated the treatment effects contradictorily, with three showing a reduced risk of the outcome, two showing a null effect, and one showing an increased risk. These results demonstrate that different time-zero settings can lead to different conclusions, even if the same dataset is analyzed for the same research question, shedding light on the significant impact of time-zero settings on the effect estimates in such a comparative effectiveness study.

No "gold standard" exists for setting time zero in studies with a non-user comparator, and the present study does not aim to determine superiority among the six setting methods. Nevertheless, SED versus SED (cloning method) and TI versus Matched (systematic order) were based on approaches proposed to avoid possible biases, such as

Table 1 Baseline characteristics of patients in each group according to time zero setting methods

	Treatment group		Non-use group				
	SED (<i>n</i> = 10,929)	TI (<i>n</i> = 10,929)	SED (vs SED naïve approach) (<i>n</i> = 24,767)	SED (vs TI) (<i>n</i> = 23,935)	Random (<i>n</i> = 21,860)	Matched (random order) (<i>n</i> = 10,929)	Matched (systematic order) (<i>n</i> = 10929)
Men, <i>n</i> (%)	7985 (73.1)	7985 (73.1)	17942 (72.4)	17374 (72.6)	15891 (72.7)	8007 (73.3)	8030 (73.5)
Age, years, mean (SD)	52.4 (9.5)	52.4 (9.5)	51.6 (10.8)	51.6 (10.8)	51.7 (10.9)	52.4 (9.5)	52.3 (9.5)
Type 2 diabetes duration, years, mean (SD)	1.1 (1.7)	1.1 (1.7)	1.0 (1.7)	1.0 (1.7)	1.0 (1.7)	1.0 (1.7)	1.1 (1.7)
Duration from SED to time zero, days, mean (SD)	0.0 (0.0)	154.4 (345.5)	0.0 (0.0)	0.0 (0.0)	306.6 (435.0)	154.4 (345.5)	154.4 (345.5)
Comorbidities, <i>n</i> (%)							
Cerebrovascular disease	675 (6.2)	675 (6.2)	1131 (4.6)	1088 (4.5)	1021 (4.7)	446 (4.1)	437 (4.0)
Chronic pulmonary disease	740 (6.8)	740 (6.8)	1914 (7.7)	1866 (7.8)	1766 (8.1)	751 (6.9)	785 (7.2)
Congestive heart failure	579 (5.3)	579 (5.3)	1090 (4.4)	1067 (4.5)	1007 (4.6)	359 (3.3)	364 (3.3)
Dementia	19 (0.2)	19 (0.2)	46 (0.2)	46 (0.2)	44 (0.2)	19 (0.2)	19 (0.2)
Diabetes with chronic complication	1042 (9.5)	1042 (9.5)	2000 (8.1)	1852 (7.7)	1569 (7.2)	945 (8.6)	940 (8.6)
Diabetes without chronic complication	759 (6.9)	759 (6.9)	2309 (9.3)	2209 (9.2)	2003 (9.2)	963 (8.8)	929 (8.5)
Mild liver disease	2383 (21.8)	2383 (21.8)	5114 (20.6)	4972 (20.8)	4579 (20.9)	2264 (20.7)	2257 (20.7)
Myocardial infarction	251 (2.3)	251 (2.3)	94 (0.4)	90 (0.4)	84 (0.4)	37 (0.3)	36 (0.3)
Peptic ulcer disease	786 (7.2)	786 (7.2)	1990 (8.0)	1934 (8.1)	1788 (8.2)	836 (7.6)	827 (7.6)
Peripheral vascular disease	552 (5.1)	552 (5.1)	835 (3.4)	801 (3.3)	732 (3.3)	376 (3.4)	382 (3.5)
Rheumatic disease	119 (1.1)	119 (1.1)	333 (1.3)	326 (1.4)	309 (1.4)	140 (1.3)	141 (1.3)
Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin	257 (2.4)	257 (2.4)	1490 (6.0)	1465 (6.1)	1385 (6.3)	531 (4.9)	524 (4.8)
Hemiplegia or paraplegia	48 (0.4)	48 (0.4)	122 (0.5)	117 (0.5)	105 (0.5)	32 (0.3)	31 (0.3)
Renal disease	136 (1.2)	136 (1.2)	336 (1.4)	326 (1.4)	299 (1.4)	111 (1.0)	105 (1.0)
Moderate or severe liver disease	5 (0.0)	5 (0.0)	121 (0.5)	121 (0.5)	114 (0.5)	26 (0.2)	26 (0.2)
AIDS/HIV	3 (0.0)	3 (0.0)	14 (0.1)	13 (0.1)	13 (0.1)	4 (0.0)	4 (0.0)
Metastatic solid tumor	28 (0.3)	28 (0.3)	345 (1.4)	342 (1.4)	332 (1.5)	37 (0.3)	37 (0.3)
Antihypertensive medications, <i>n</i> (%)	1709 (15.6)	1709 (15.6)	6512 (26.3)	6403 (26.8)	6051 (27.7)	1714 (15.7)	1713 (15.7)
Anticoagulant agents, <i>n</i> (%)	557 (5.1)	557 (5.1)	1831 (7.4)	1791 (7.5)	1697 (7.8)	525 (4.8)	521 (4.8)
Follow-up period, days, mean (SD)	947.4 (763.3)	793.0 (703.9)	627.7 (643.5)	639.8 (645.5)	366.7 (439.2)	536.9 (567.8)	666.6 (639.6)

Table 1 (continued)

Data are *n* (%) or mean (standard deviation)

Treatment group: SED: treated patients in the method of SED versus SED (naïve approach), TI: treated patients in the method of TI versus SED, TI versus Random, TI versus Matched (random order), and TI versus Matched (systematic order)

Non-use group: SED (naïve approach); non-user patients in the method of SED versus SED (naïve approach), SED (vs TI): non-user patients in the method of TI versus SED, Random: non-user patients in the method of TI versus Random, Matched (random order): non-user patients in the method of TI versus Matched (random order), Matched (systematic order): non-user patients in the method of TI versus Matched (systematic order)

For SED versus SED (cloning method), the baseline characteristics of the patients are not shown because the method used an artificial population, and patient characteristics were the same in both groups because of the methodology

SD standard deviation, SED study entry date (the date of the first prescription of a glucose-lowering agent), TI treatment initiation (first prescription date of lipid-lowering agents)

Table 2 Incidence rate, rate ratio, and cumulative incidence of the outcome in each time zero setting method

Time-zero setting method	Group	Number of patients	Number of events	Person-years	Incidence rate per 1000 person-year (95% CI)	Rate ratio (95% CI)	Cumulative incidence at 1 year follow-up (95% CI)
SED vs SED (naïve approach)	Non-use	24,767	4040	42,590	94.9 (91.9–97.8)	–	0.124 (0.119–0.128)
	Treatment	10,929	1641	28,367	57.8 (55.0–60.6)	0.61 (0.58–0.65)	0.070 (0.065–0.075)
SED vs SED (cloning method) ^a	Non-use	–	–	–	–	–	0.100 (0.096–0.103)
	Treatment	–	–	–	–	–	0.101 (0.097–0.104)
TI vs SED	Non-use	23,935	3208	41,952	76.5 (73.8–79.1)	–	0.098 (0.094–0.103)
	Treatment	10,929	1641	23,746	69.1 (65.8–72.5)	0.90 (0.85–0.96)	0.087 (0.081–0.092)
TI vs Random	Non-use	21,860	1133	21,960	51.6 (48.6–54.6)	–	0.054 (0.050–0.058)
	Treatment	10,929	1641	23,746	69.1 (65.8–72.5)	1.34 (1.24–1.44)	0.087 (0.081–0.092)
TI vs Matched (random order)	Non-use	10,929	1620	16,076	100.8 (95.9–105.7)	–	0.122 (0.115–0.129)
	Treatment	10,929	1641	23,746	69.1 (65.8–72.5)	0.69 (0.64–0.73)	0.087 (0.081–0.092)
TI vs Matched (systematic order)	Non-use	10,929	1441	19,961	72.2 (68.5–75.9)	–	0.090 (0.084–0.096)
	Treatment	10,929	1641	23,746	69.1 (65.8–72.5)	0.96 (0.89–1.03)	0.087 (0.081–0.092)

SED versus SED (naïve approach): time zero set at SED for both groups (naïve approach)

SED versus SED (cloning method): time zero set at SED for both groups (cloning method)

TI versus SED: time zero set at treatment initiation for the treatment group and SED for the non-use group

TI versus Random: time zero set at treatment initiation for the treatment group and at a randomly sampled hospital visit date for the non-use group

TI versus Matched (random order): time zero set at treatment initiation for the treatment group and at the matched date for the non-use group (matching in random order)

Treatment initiation versus Matched (systematic order): time zero set at treatment initiation for the treatment group and at the matched date for the non-use group (matching in systematic order)

CI confidence interval, HR hazard ratio, SED study entry date (the date of the first prescription of a glucose-lowering agent plus the diagnostic record of type 2 diabetes in the same month), TI treatment initiation (first prescription date of lipid-lowering agents)

^aThe calculation of incidence rate and crude incidence rate ratio was not applicable in SED versus SED (cloning method), due to the use of an artificial population. Thus, only the cumulative incidence at 1 year follow-up was estimated

immortal-time bias, time-lag bias, and selection bias [11, 25, 28]. Therefore, these two settings would presumably have higher between-group comparability than the other settings. For example, SED versus SED (cloning method) used the method proposed by Hernán et al [11, 25]. By allocating clones of each patient to the respective arms,

the patients in both arms became identical at time zero, eliminating confounding at baseline. Moreover, despite the time zero set at SED, clones and artificial censoring in this method successfully avoided immortal-time bias, which would otherwise occur when time zero is set at SED in both groups [9].

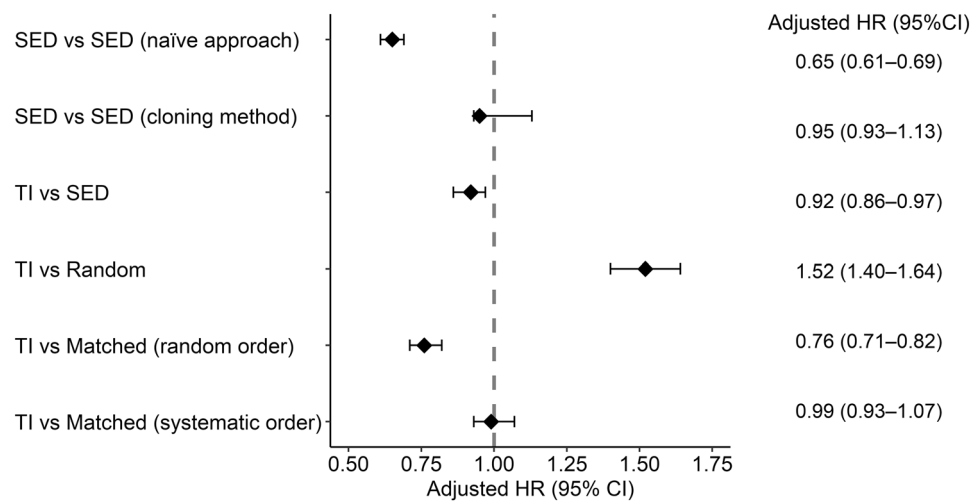


Fig. 2 Hazard ratio in each time-zero setting method. SED versus SED (naïve approach): time zero set at SED for both groups (naïve approach). SED versus SED (cloning method): time zero set at SED for both groups (cloning method). TI versus SED: time zero set at treatment initiation for the treatment group and SED for the non-use group. TI versus Random: time zero set at treatment initiation for the treatment group and at a randomly sampled hospital visit date for the non-use group. TI versus Matched (random order): time zero set at treatment initiation for the treatment group and at the matched date for the non-use group (matching in random order). TI versus Matched

(systematic order): time zero set at treatment initiation for the treatment group and at the matched date for the non-use group (matching in systematic order). For SED versus SED (cloning method), only the “group” was included as an explanatory variable in the Cox regression model because covariate adjustment was already performed in the estimation of weights for the inverse-probability-of-censoring weighting. *CI* confidence interval, *HR* hazard ratio, *SED* study entry date (the date of the first prescription of a glucose-lowering agent plus the diagnostic record of type 2 diabetes in the same month), *TI* treatment initiation (first prescription date of lipid-lowering agents)

In TI versus Matched (systematic order), the duration from SED to time zero was equated in the two groups by matching, preventing time-lag bias. Indeed, this time-zero setting has been reported as a good setting example for non-users [12]. However, the way of choosing non-users can introduce a selection bias [28]. For example, in random order matching as in TI versus Matched (random order), a non-user, who should be excluded due to an outcome event before time zero if matched with a particular patient, may be included in the cohort by being matched with another treatment patient with a different time zero. In other words, a non-user patient who should be excluded because of the outcome event remains in the pool of potential matches of subsequent patients in the treatment group. As a result, the probability of being matched for a non-user who had outcome events is increased and more patients with outcomes would be included in the non-use group [8, 28]. One way to avoid this bias is to select non-users chronologically or systematically such that a non-user who was once discarded will never be included as a match for another treatment patient [28]. In this sense, TI versus Matched (systematic order) likely managed to avoid this selection bias by systematically matching patients based on the length from SED to time zero.

As such, SED versus SED (cloning method) and TI versus Matched (systematic order) seemed to control for various biases, possibly preserving high between-group

comparability. Interestingly, both methods yielded an adjusted HR of approximately 1, indicating a null effect of lipid-lowering treatment for DR. There is no known “true value” to be compared against; however, previous randomized clinical trial (RCT) showed no difference in progression of DR between lipid lowering treatment and placebo group in patients without pre-existing DR [29], although a risk reduction was shown in patients with pre-existing DR [29, 30]. Since the study population of this current study comprised patients without previous DR, the results of the above two methods were similar to RCT findings. As explained above, the results of these two methods would have been much less biased than the results obtained by other setting methods, which probably had lower group comparability. Therefore, we discuss the results of the remaining four settings in terms of potential biases and their directions.

Three methods (SED vs SED [naïve approach], TI versus SED, and TI versus Matched [random order]) showed an adjusted HR of < 1, indicating that the treatment had a protective effect on the study outcome. However, these reduced outcome risks were likely attributed to the overestimation of the incidence rate in the non-use group due to biases caused by the respective time-zero settings. First, SED versus SED (naïve approach) is a typical case of immortal-time bias [7]. This may have overestimated the incidence rate in the non-use group, causing a downward bias in the relative risk in the treatment group. Second,

in TI versus SED, the follow-up period of the non-use group included the early days after SED, during which outcome events occurred more frequently, whereas those of the treatment group did not, which is a case of bias from time lag and latency [8]. This probably resulted in overestimating the incidence rate in the non-use group. The disproportionately high incidence rate soon after SED may be because patients were encouraged to undergo an ophthalmologic examination at diagnosis [31] or did so driven by the start of pharmacotherapy. After excluding patients with outcomes in the early days after time zero as a sensitivity analysis, thereby removing the imbalance associated with this period, TI versus SED no longer showed a treatment effect (Online Resource Table S1). Finally, TI versus Matched (random order) probably introduced a selection bias by matching non-users in random order as explained earlier [8].

In contrast, TI versus Random yielded an adjusted HR of >1 , indicating that treatment had an increased outcome risk. In this method, time zero of non-users was randomly selected from all hospital visit dates after SED. Although this type of solution was proposed in the previous literature [11], our results suggested that this was not suitable in this study. In particular, this setting would introduce a selection bias by removing those with an event before time zero in the non-use group, which had longer period from SED to time zero (154.4 days in the treatment group vs 306.6 days in the non-use group).

To the best of our knowledge, this is the first study to demonstrate the direction and magnitude of potential biases caused by different time-zero settings in a non-user comparative effectiveness study using RWD. We assumed various time-zero settings based on previous studies, such as those considered problematic or desirable [8, 11, 12, 25, 28]. A major strength was that we applied six possible time-zero settings to the same dataset for the same research question using real-world data, purely focusing on evaluating the impact of time-zero settings.

However, the present study had a limitation in that there may have been residual biases in our effect estimates due to, for example, confounding by indication and time-varying confounding [28]. The covariate information used for the cloning method, calculation of propensity score, and Cox regression model for outcome were based on baseline data only, which may lead to residual bias due to time varying confounding. Also, we did not use analytical methods that can control for immortal time biases, such as the time-dependent Cox regression model. For TI versus Random method, although there might be more suitable approaches, such as selecting visits randomly based on the distribution of times in the treatment group, these were not applied. Nevertheless, this limitation does not degrade the value of this study because our aim was not to obtain accurate effect estimates. For simplicity, the present study focused on time-zero settings without considering other aspects, including latency

time [5], grace period [14], and treatment discontinuation [1]. An analysis of the impact of these other aspects of the study design is warranted in future studies. It should also be noted that the present study did not aim to determine superiority among different time-zero settings. A desirable time-zero setting should vary depending on the study. Thus, the setting used for a particular study should be individually justified, considering the study design, analytical methods, diseases of interest, outcomes, and exposures. Finally, only a single data source was used in this study, which may limit the generalizability of these findings. However, the used data source contains large numbers of patients and is quite representative of the active population in Japan.

5 Conclusion

In conclusion, the results of this study illustrated that the conclusions regarding treatment effect varied depending on how to set time zero in a non-user comparative effectiveness study using RWD, highlighting the significant impacts of time-zero settings on the study results. An improper setting of time zero may lead to biased and misleading conclusions. Therefore, to minimize biases, researchers should carefully select time zero to increase between-group comparability when designing a study. Further research on the impact of time-zero settings in studies of different diseases, outcomes, and exposures will facilitate further understanding of this important element that should be considered when designing a comparative effectiveness study using RWD.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40801-022-00343-1>.

Acknowledgments Clinical Study Support, Inc provided the medical writing support.

Declarations

Funding This study was conducted as part of a joint research course (Real-World Evidence and Data Assessment) between Juntendo University and Shin Nippon Biomedical Laboratories Ltd.

Conflict of Interest RW, TH, and TL are employees of Clinical Study Support, Inc.

Data Availability The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics Approval The ethics committee of Juntendo University, Tokyo, Japan, approved this study (approval number: E21-0041). This study was conducted in accordance with the provisions of the Declaration of Helsinki.

Consent to Participate This observational study used only secondary, fully anonymized data. An observational study that exclusively uses anonymized data is outside the scope of the Ethical Guidelines

for Medical and Health Research Involving Human Subjects in Japan. Therefore, informed consent was not required for this study.

Author Contributions RW and TH contributed to the study conception and design. TH and TL conducted the statistical analyses. RW contributed to the manuscript drafting. All authors contributed to data interpretation, revised the manuscript critically for important intellectual content, and approved the final manuscript for submission. All authors read and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Schneeweiss S, Patorno E. Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. *Endocr Rev*. 2021;42:658–90.
- Schneeweiss S. Improving therapeutic effectiveness and safety through big healthcare data. *Clin Pharmacol Ther*. 2016;99:262–5.
- Nishioka K, Makimura T, Ishiguro A, et al. Evolving acceptance and use of RWE for regulatory decision making on the benefit/risk assessment of a drug in Japan. *Clin Pharmacol Ther*. 2022;111:35–43.
- Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58:323–37.
- Suissa S, Dell’Aniello S. Time-related biases in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 2020;29:1101–10.
- Iudici M, Porcher R, Riveros C, et al. Time-dependent biases in observational studies of comparative effectiveness research in rheumatology. A methodological review. *Ann Rheum Dis*. 2019;78:562–9.
- Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol*. 2008;167:492–9.
- Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care*. 2012;35:2665–73.
- Lévesque LE, Hanley JA, Kezouh A, et al. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ*. 2010. <https://doi.org/10.1136/bmj.b5087>.
- van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol*. 2012;65:126–31.
- Hernán MA, Sauer BC, Hernández-Díaz S, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70–5.
- Luijken K, Spekrijse JJ, van Smeden M, et al. New-user and prevalent-user designs and the definition of study time origin in pharmacoepidemiology: a review of reporting practices. *Pharmacoepidemiol Drug Saf*. 2021;30:960–74.
- Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol*. 2015;11:437–41.
- Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183:758–64.
- Yang X, So WY, Ma RCW, et al. Low HDL cholesterol, metformin use, and cancer risk in type 2 diabetes. *Diabetes Care*. 2011;34:375–80.
- Kobayashi T, Uda A, Udagawa E, et al. Lack of increased risk of lymphoma by thiopurines or biologics in Japanese patients with inflammatory bowel disease: a large-scale administrative database analysis. *J Crohns Colitis*. 2020;14:617–23.
- Yamazaki K, Macaulay D, Song Y, et al. Clinical and economic burden of patients with chronic hepatitis C with versus without antiviral treatment in Japan: an observational cohort study using hospital claims data. *Infect Dis Ther*. 2019;8:285–99.
- Korol S, White M, O’Meara E, et al. Is there a potential association between spironolactone and the risk of new-onset diabetes in a cohort of older patients with heart failure? *Eur J Clin Pharmacol*. 2019;75:837–47.
- Hatswell AJ, Deighton K, Snider JT, et al. Approaches to selecting “time zero” in external control arms with multiple potential entry points: A simulation study of 8 approaches. *Med Decis Mak*. 2022. <https://doi.org/10.1177/0272989X221096070>.
- Backenroth D. How to choose a time zero for patients in external control arms. *Pharm Stat*. 2021;20:783–92.
- Kawasaki R, Konta T, Nishida K. Lipid-lowering medication is associated with decreased risk of diabetic retinopathy and the need for treatment in patients with type 2 diabetes: a real-world observational analysis of a health claims database. *Diabetes Obes Metab*. 2018;20:2351–60.
- Laurent T, Simeone J, Kuwatsuru R, et al. Context and considerations for use of two Japanese real-world databases in Japan: Medical Data Vision and Japanese Medical Data Center. *Drugs Real World Outcomes*. 2022;9:175–87.
- Nagai K, Tanaka T, Kodaira N, et al. Data resource profile: JMDC claims database sourced from health insurance societies. *J Gen Fam Med*. 2021;22:118–27.
- Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–9.
- Maringe C, Benitez Majano S, Exarchakou A, et al. Reflection on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. *Int J Epidemiol*. 2020;49:1719–29.
- Willems S, Schat A, van Noorden MS, et al. Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator. *Stat Methods Med Res*. 2018;27:323–35.
- Libby G, Donnelly LA, Donnan PT, et al. New users of metformin are at low risk of incident cancer: a cohort study among people with type 2 diabetes. *Diabetes Care*. 2009;32:1620–5.
- Suissa S, Moodie EE, Dell’Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf*. 2017;26:459–68.
- Keech AC, Mitchell P, Summanen PA, et al. Effect of fenofibrate on the need for laser treatment for diabetic retinopathy (FIELD study): a randomised controlled trial. *Lancet*. 2007;370:1687–97.
- Shi R, Zhao L, Wang F, et al. Effects of lipid-lowering agents on diabetic retinopathy: a meta-analysis and systematic review. *Int J Ophthalmol*. 2018;11:287–95.
- The Japan Diabetes Society. Japanese clinical practice guideline for diabetes 2019. Tokyo: Nankodo; 2019.