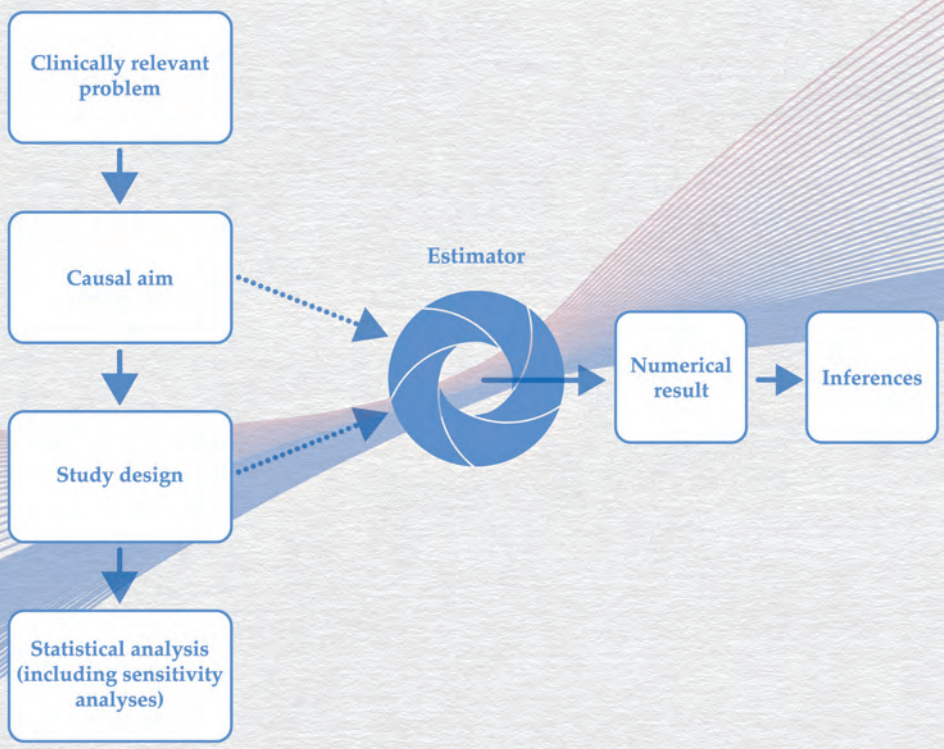
The impact of epidemiologic methods on findings in studies of causal effects and prediction modelling

Luijken, K.

2


Clinically relevant problem
Causal aim
Study design
Statistical analysis (including sensitivity analyses)
Estimator
Numerical result
Inferences

New-user and prevalent-user designs and the definition of study time origin in pharmacoepidemiology: a review of reporting practices

Guidance reports for observational comparative effectiveness and drug safety research recommend implementing a new-user design whenever possible, since it reduces the risk of selection bias in exposure effect estimation compared to a prevalent-user design. The uptake of this guidance has not been studied extensively. We reviewed 89 observational effectiveness and safety cohort studies published in six pharmacoepidemiologic journals in 2018 and 2019. We developed an extraction tool to assess how frequently new-user and prevalent-user designs were reported to be implemented. For studies that implemented a new-user design in both exposure arms, we extracted information about the extent to which the moment of meeting eligibility criteria, treatment initiation, and start of follow-up were reported to be aligned. Of the 89 studies included, 40% reported implementing a new-user design for both the study exposure arm and the comparator arm, while 13% reported implementing a prevalent-user design in both arms. The moment of meeting eligibility criteria, treatment initiation, and start of follow-up were reported to be aligned in both exposure arms in 53% of studies that reported implementing a new-user design. We provided examples of studies that minimized the risk of introducing bias due to unclear definition of time origin in unexposed participants, immortal time, or a time lag. To sum up, almost half of the included studies reported to implement a new-user design. Implications of misalignment of study design origin were difficult to assess because it would require explicit reporting of the target causal effect in original studies. We recommend that the choice for a particular study time origin is explicitly motivated to enable assessment of validity of the study.

This chapter was based on: Luijken K, Spekreijse JJ, van Smeden M, Gardarsdottir H, Groenwold RHH. New-user and prevalent-user designs and the definition of study time origin in pharmacoepidemiology: A review of reporting practices. *Pharmacoepidemiology and Drug Safety*. 2020;30(7):960-74.

1 | Background

Guidance reports for comparative effectiveness and safety research of pharmacological treatments recommend the new-user design[1-4], in which follow-up time generally starts with the first prescription or dispensing of the drug(s) of interest[5]. In contrast, in the prevalent-user design both current (prevalent) and new users of a drug are included. The new-user design enforces appropriate temporal ordering of measurements of confounders, treatment, and outcome, protecting the researcher against accidental adjustment for variables affected by treatment and against finding associations that are based on reversed causation[1-8]. However, the start of a treatment can be difficult to capture (especially in case of intermittently used treatments) and exclusion of prevalent users may reduce follow-up time or sample size[5,7-10]. It is unclear how often and for which reasons researchers deviated from the guidance to implement a new-user design.

To assess the uptake of new-user design guidance, it is important to go beyond the distinction of including new or prevalent users. Many time-related biases can be prevented by choosing a study time origin (or study baseline) such that it establishes alignment of the moment of meeting eligibility criteria, treatment initiation, and start of follow-up[6,11-13]. Previous studies investigated how often pharmacoepidemiologic studies deviated from the recommendation to implement a new-user design[14-16], however, the implementation of new-user designs in terms of alignment of eligibility, treatment initiation, and start of follow-up has not been studied yet.

In the current study, we reviewed the literature about contemporary observational effectiveness and safety cohort studies. We assessed how frequently new-user and prevalent-user designs were reported to be implemented in studies published in high-ranked pharmacoepidemiologic journals. For studies implementing a new-user design, we evaluated to what extent eligibility, treatment initiation, and start of follow-up were reported to be aligned.

2 | Defining a study time origin

Here, we briefly review common biases that can be introduced by inappropriate designation of follow-up time in observational studies (for a more elaborate discussion, see e.g.,[3,5-7,11,12,17-19]). Ideally, a study is designed such that operational decisions match the specified causal contrast of interest, i.e., the target causal effect or so-called estimand. Whether the target causal effect can be estimated from the available data, depends on

whether it is convincing that untestable identifying assumptions hold, i.e., conditional exchangeability, positivity and consistency. The choice of the time origin of a study design is directly linked to the target causal effect, because the estimated outcome risk refers to the (cumulative) probability of an event of interest occurring over time since a given origin in a specific population[18].

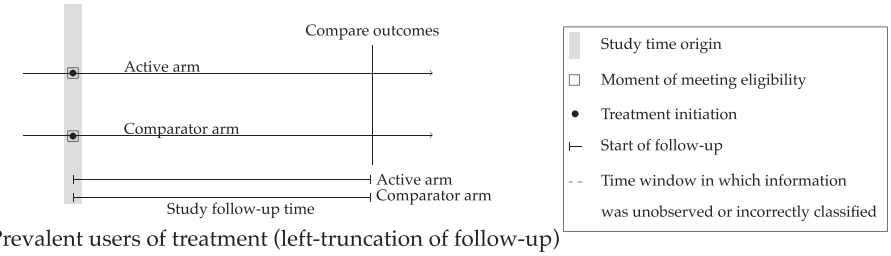
Figure 1 shows a simplified version of a study of a binary treatment in which eligibility, treatment initiation, and start of follow-up correspond[4]. An archetypical study design is the active-comparator incident-user design, that potentially allows identification of the target causal effect from empirical data. When an appropriate active comparator is chosen, in the sense that the comparison group reflects a clinically meaningful alternative treatment option in real-world practice, the active-comparator incident-user design increases the likelihood of achieving conditional exchangeability by minimizing the risk of confounding by indication and selection bias[20].

When prevalent users of treatment are included in a study, the follow-up of those individuals is left truncated and omitted from the analysis. For permanent outcomes, of which 'death' is arguably the clearest example, individuals included in the analysis did not develop the outcome during the exposed period before start of follow-up. Consequently, including prevalent users can lead to under-ascertainment of events early in the course of treatment and increases the risk of controlling for confounding factors that were affected by the treatment[5,21]. Since much information is unobserved, it will be difficult to specify the set of covariates that are sufficient to achieve conditional exchangeability of treatment groups. In many cases, it is unlikely that a prevalent-user study design is suitable to identify the specified target causal effect (Figure 1b).

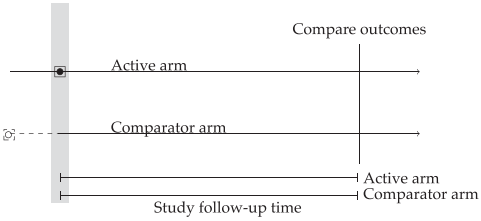
Another issue is that individuals could be assigned to a treatment group based on the treatment strategy observed after start of follow-up, rather than the treatment strategy at the time of start of follow-up[11] (Figure 1c), for example when individuals are classified as 'users' only after they filled a number (e.g., three) of prescriptions of that treatment. In that case, individuals cannot experience the outcome during the first three prescriptions. This period is often referred to as *immortal time*. When some individuals are immortal for part of the time they were followed-up, it seems unlikely that the target causal effect can be identified.

A final example of a study design that can impede identification of the target causal effect is when a time-lag bias is introduced. This happens when follow-up is started at the moment of treatment initiation, but the compared treatments are prescribed at different stages of the disease[13]. For instance, if the effect of a second-line drug

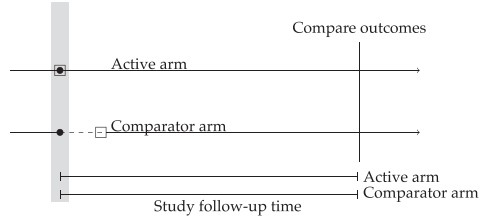
(a) Ideal causal contrast



(b) Prevalent users of treatment (left-truncation of follow-up)



(c) Follow-up time incorrectly allocated (immortal time)



(d) Time-lag in start of follow-up (inexchangeability)

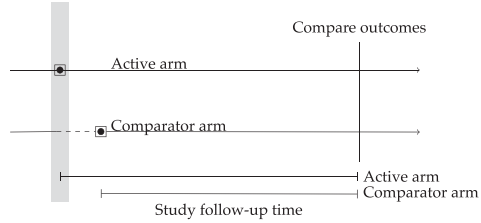


Figure 1. Schematic depiction of possible operationalizations of a study time origin. Subfigure (a) depicts a simplified ideal causal contrast for a binary treatment. Subfigures (b) - (d) depict possible biases that can be introduced by inappropriate designation of follow-up time. In empirical studies, the specified design flaws can occur in either or both exposure arms and a combination of flaws can occur. For simplification, we presented a single design flaw in the comparator exposure arm in each subfigure (b) - (d). Study design flaws may lead to violation of identifying assumptions, as is explained in Section 2 of the main text. The dotted lines in the figure thus indicate the consideration whether exchangeability of treatment groups is jeopardized by the misalignment and whether this can be corrected by measured covariates, or non-positivity is introduced.

is compared to a first-line drug and the start of follow-up for both the active and comparator arm is defined by treatment initiation, the disease stage differs between the treatment groups. As capturing this difference in disease progress in measured covariates is hardly feasible[22], time-lag bias likely jeopardises the (conditional) exchangeability of treatment groups (Figure 1d). When protocol adherence to switch to second line is high and precise, the incomparability of treatment groups may be so extreme that non-positivity is introduced.

3 | Methods

We systematically assessed the reporting practices in observational studies of treatment effects regarding the definition of the study time origin and inclusion of new versus prevalent users of treatment. A protocol of this study is available on Open Science Framework[23]. Based on recommendations by the editor and reviewers, we deviated from this protocol. Specifically, while we scored the items of the extraction tool for all included articles, we discuss the results on alignment in study design origin for new-user designs only, as will be explained below. This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[24], where applicable.

3.1 | Journal selection and included type of studies

We aimed to review the reporting of approximately 100 articles published before the 1st of July 2019 in journals publishing pharmacoepidemiologic studies of drug-outcome associations. Six pharmacoepidemiologic journals were included: Annals of Pharmacotherapy, British Journal of Clinical Pharmacology, Drug Safety, European Journal of Clinical Pharmacology, Pharmacotherapy, and Pharmacoepidemiology and Drug Safety. These state-of-the art pharmacoepidemiologic journals were selected because reporting on study design implementation was expected to be relatively complete. We performed a PubMed search on February 3rd 2020 (see protocol[23] for search string) which returned 2,457 records. Study inclusion criteria were: study described original pharmacoepidemiologic research into the relation between drug exposure and a clinical outcome; data were collected for research purposes or obtained from routinely collected health data; data were gathered according to a cohort study design, since the definition of new versus prevalent users is not as straightforward in other designs, such as a cross-sectional, case-crossover or case-control design. Exclusion

criteria were: pharmacokinetic-pharmacodynamic studies; cost-effectiveness studies; data on treatment exposure were collected through self-report. We also excluded studies of vaccination, antibiotic treatment of a single treatment episode (up to 10 days), chemotherapy, or intravenous drugs, because for these kinds of interventions new-user designs are more natural. KL screened the title and abstract of all studies that result from the searches and included relevant articles based on the eligibility criteria. We applied a quota sampling strategy[25] and continued screening articles until we reached the most recent 100 articles published before July 1st, 2019.

3.2 | Extraction of study characteristics and evaluation of reporting quality

Articles were scored on a set of items derived from guideline recommendations about elements that should be reported in protocols[20,26] or articles[4,27] of effectiveness and safety research using large observational databases, as well as methodological articles that discuss the study time origin in observational studies of causal effects[6,11]. The main focus was on the distinction between new-user and prevalent-user designs and the alignment of moment of meeting eligibility criteria, moment of treatment initiation, and start of follow-up in new-user designs. The established scoring tool was pilot tested on six randomly chosen included studies by KL and JS and further adjusted (all items can be found in Table 2 and 3).

An incident user can more generally be defined as a new user of any treatment decision, i.e., initiating a treatment, but also switching to a different treatment or a change of dose. This understanding of the incident-user design was introduced by Brookhart[28] and expanded in work by Suissa[29] and will be used throughout the current study. For the item that scored reporting of whether the comparator exposure arm implemented a new-user or prevalent-user design, we decided to score nonusers of treatment as prevalent users. Whereas non-use is not associated with the biases typically associated with prevalent users (e.g., adjusting for intermediates, depletion of susceptibles), definition of study time origin in studies with a non-user comparator arm is complicated because the choice of the time origin since which the (cumulative) probability of an event of interest can occur in the specified population may not be as straightforward for non-users of treatment. Consequently, it is more challenging to assess whether the study exposure arm and comparator arm can be assumed to be comparable conditional on measured confounders (i.e., whether there is conditional exchangeability).

Information was gathered on general characteristics of the included studies; funding source, type of data source, patient domain, sample size, and length of enrollment window. Funding source was defined as 'private' when the article stated the study

was funded by a pharmaceutical company or when any of the authors was affiliated with a pharmaceutical company and defined as 'public' otherwise. Data sources were classified into hospital data, dispensings, prescriptions, or claims. Patient domain was grouped into medical specialties based on the target population that was mentioned in the article objective. When the target population did not match a single medical specialty, information on the type of treatment and study outcome was used to identify the medical specialty.

Articles were reviewed independently by KL and JS, results were discussed between the two reviewers and in case of disagreement a third reviewer (RG) was consulted. When multiple effectiveness or safety analyses were described in a single article, only the first-reported analysis was scored. When subgroup analyses were performed in the included studies, only the main analysis was scored. When methods were discussed in an online protocol or described in a different article, we reviewed the referred material.

3.3 | Data synthesis

Rater agreement was computed using the unweighted Cohen's kappa for nominal variables and two coders[30]. Cohen's kappa ranges from -1 (perfect disagreement) to 1 (perfect agreement). Reporting of items was presented as percentages of total number of included studies and 95% confidence intervals (CIs) were computed using the normal approximation.

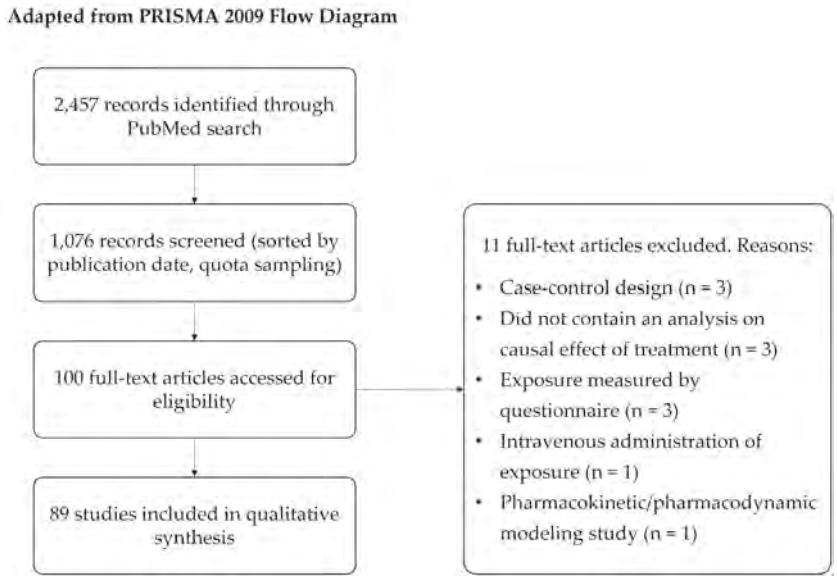


Figure 2. The screening and inclusion of eligible articles.

4 | Results

After screening the full texts of the 100 articles included during abstract and title screening, 89 studies remained based on the eligibility criteria (see Figure 2). The characteristics of the 89 included studies are summarized in Table 1. The most common patient domains considered were cardiology (17 %), neurology (11%) and primary care (10%). The median sample size was 7,011 (range 14 - 3,351,674). In 10% of studies (n = 9), a sample size calculation was reported. The length of follow-up ranged from 1 hour follow-up in one study to a median follow-up of 13.6 years in another study. Rater agreement is presented in Figure 3. Item kappas indicated that agreement between raters was low (range 0.05–0.75), which was mostly due to ambiguous reporting of the extracted information. Despite the low rater agreement of the initial scores, the presented results have a meaningful interpretation since consensus was reached for all scores with initial disagreement.

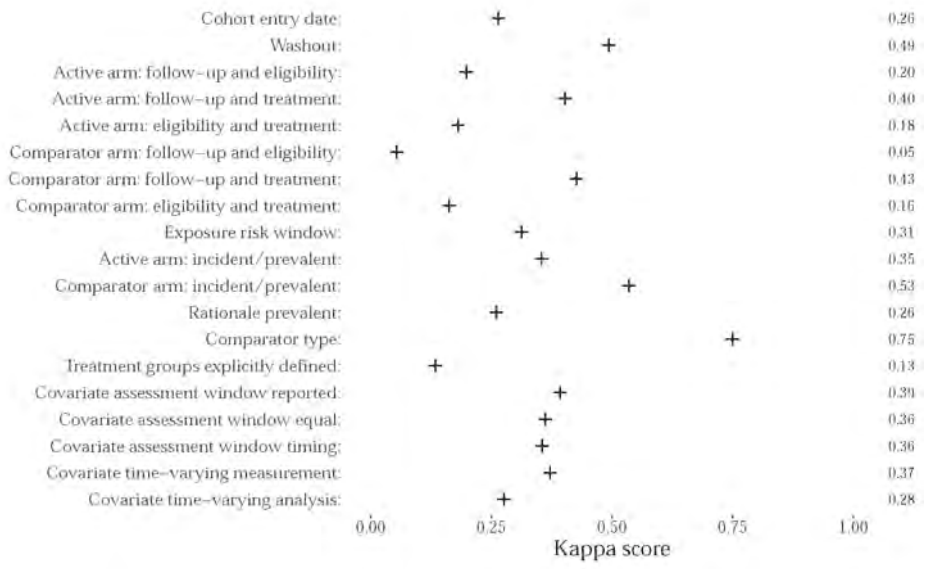


Figure 3. Agreement between raters, measured by Cohen's kappa (unweighted).

Table 1 Characteristics of the 89 included studies.

Item	Item options	Number of studies (proportion)
Journal	Annals of Pharmacotherapy	16 (0 .18)
	British Journal of Clinical Pharmacology	12 (0 .13)
	Drug Safety	11 (0 .12)
	European Journal of Clinical Pharmacology	8 (0 .09)
	Pharmacoepidemiology and Drug Safety	27 (0 .30)
	Pharmacotherapy	15 (0 .17)
Continent	Africa	1 (0.01)
	Asia	16 (0.18)
	Europe	30 (0.34)
	North America	37 (0.42)
	Oceania	2 (0.02)
	Multiple	1 (0.01)
	Not reported	2 (0.02)
Year of publication	2018	56 (0.63)
	2019	33 (0.27)
Funding	Non-pharmaceutical	83 (0.93)
	Pharmaceutical	6 (0.07)
Data source type	Claims	32 (0.36)
	Dispensing	19 (0.21)
	Hospital data	26 (0.29)
	Prescription	11 (0.12)
	Dispensing and prescription	1 (0.01)
Domain	Cardiology	15 (0.17)
	Neurology	10 (0.11)
	Primary care	9 (0.10)
	Infectious disease	6 (0.07)
	Nephrology	6 (0.07)
	Other	43 (0.48)
Sample size	< 500	23 (0.26)
	500 – 50,000	44 (0.49)
	> 50,000	22 (0.25)
Sample size calculation	No	80 (0.90)
	Yes	9 (0.10)
If sample size calculation reported, size reached?	No	1 (0.11)
	Yes	7 (0.78)
	Unclear	1 (0.11)
Cohort entry[10]	Event-based	22 (0.25)
	Exposure-based	28 (0.31)
	Multiple event-based	33 (0.37)
	Time-based	6 (0.07)
Study entry level [3, item C2]	Episode	6 (0.07)
	Person	83 (0.93)

4.1 | New-user and prevalent-user designs

An overview of item scores is given in Table 2. Forty percent of studies (95% CI 30% - 51%, n = 36) reported implementing a new-user design for both the study exposure arm and the comparator exposure arm, while 13% (7% - 22%, n = 12) reported implementing a prevalent-user design for both exposure arms (Figure 4). In 58% (42% - 74%, n =21) of studies with a new-user design for both exposure arms a washout for exposure was reported. For 6% of studies (1% - 10%, n = 5) it was unclear whether a new-user or a prevalent-user design was implemented. When a prevalent-user design was reported to be implemented, three studies provided a rationale for including prevalent users. The motivation to include prevalent users concerned biological plausibility of a cumulative effect on outcome risk[31-33].

4.2 | Alignment in new-user designs

In the 36 studies that reported implementing a new-user design in both exposure arms, moment of meeting eligibility criteria, treatment initiation, and start of follow-up were reported to be aligned in both exposure arms in 53% of studies (36% - 69%, n = 19). Moment of meeting eligibility criteria, start of treatment, and start of follow-up were reported to be misaligned in both exposure arms in 6% of studies (0% - 13%, n = 2) and alignment was unclear in 6% of studies (0% - 13%, n = 2) (Figure 3). In the remaining studies (n = 13), at least one of the six alignment items was misaligned or unclear (see Table 3 for the alignment items).

Implications of misalignment of eligibility, treatment initiation, and start of follow-up can only be assessed relative to the specified target causal effect. Initially, the protocol of this study contained an item to extract whether the target causal effect was reported, but we adjusted this during the pilot phase of our extraction tool when we discovered that no study explicitly reported an estimand (see protocol revision[23] from version 2 to version 3). Based on recommendations by the editor and reviewers, we scored whether an explicit description of the target causal effect was provided in the 36 new-user active-comparator studies. Twenty-two percent of studies (9% - 36%, n = 8) provided an explicit definition of the target causal effect. In studies that did not explicitly report the target causal effect, it was often unclear which treatment strategies were compared and which treatment decision could be informed based on evidence from the conducted study.

Table 2 Summary of reporting of information extracted from 89 reviewed articles.

Item	Item options	Number of studies	Proportion (95% confidence interval)
Study exposure arm			
New/prevalent users	New users	66	0.74 (0.65; 0.83)
	Prevalent users	14	0.16 (0.08; 0.23)
	Unclear	9	0.10 (0.04; 0.16)
Comparator exposure arm			
Comparator type	Active comparator	46	0.52 (0.41; 0.62)
	Unexposed – no use	30	0.34 (0.24; 0.44)
	Unexposed – past use	3	0.03 (0.00; 0.07)
	Combination	1	0.01 (0.00; 0.03)
	Other	6	0.07 (0.02; 0.12)
	No comparator specified	3	0.03 (0.00; 0.07)
New/prevalent users	New users	38	0.43 (0.32; 0.53)
	Prevalent users	38	0.43 (0.32; 0.53)
	Unclear	5	0.06 (0.01; 0.10)
	No comparator or symmetry design	8	0.09 (0.03; 0.15)
General design features			
Treatment groups explicitly defined	Yes	84	0.94 (0.90; 0.99)
	No	5	0.06 (0.01; 0.10)
Cohort entry date reported	Yes	71	0.80 (0.71; 0.88)
	No	18	0.20 (0.12; 0.29)
Washout reported	Yes	37	0.42 (0.31; 0.52)
	No	52	0.58 (0.48; 0.69)
Exposure risk window reported	Yes	74	0.83 (0.75; 0.91)
	No	15	0.17 (0.09; 0.25)
Covariate assessment			
Covariate assessment window reported	Yes	45	0.51 (0.40; 0.61)
	No	38	0.43 (0.32; 0.53)
	Symmetry design or self-controlled	6	0.07 (0.02; 0.12)
If covariate assessment window was reported (n = 45), was the covariate assessment window equal for all covariates	Yes	20	0.44 (0.30; 0.59)
	No	24	0.53 (0.39; 0.68)
	Not reported	1	0.02 (0.00; 0.07)
If covariate assessment window was reported (n = 45), was the covariate assessment window before initiation of treatment	Yes	27	0.60 (0.46; 0.74)
	No	13	0.27 (0.14; 0.40)
	Not reported	5	0.11 (0.02; 0.20)
If exposure was time-varying (n = 18), were covariates measured time-varying	Yes	9	0.50 (0.27; 0.73)
	No	6	0.33 (0.12; 0.55)
	Not reported	3	0.16 (0.00; 0.34)
If covariates were measured time-varying (n = 12), was this incorporated in analysis	Yes	7	0.58 (0.30; 0.86)
	No	1	0.08 (0.00; 0.24)
	Not reported	4	0.33 (0.07; 0.60)

Table 3 Summary of reporting of alignment of start of follow-up, meeting eligibility criteria and treatment initiation extracted from 36 articles that implemented a new-user design in both exposure arms.

Item	Item options	Number of studies	Proportion (95% confidence interval)
Study exposure arm			
Alignment follow-up – eligibility	Yes	24	0.67 (0.51; 0.82)
	No	9	0.25 (0.11; 0.39)
	Unclear	3	0.08 (0.00; 0.17)
Alignment follow-up – treatment	Yes	26	0.72 (0.58; 0.87)
	No	3	0.08 (0.00; 0.17)
	Unclear	7	0.19 (0.07; 0.32)
Alignment eligibility – treatment	Yes	22	0.61 (0.45; 0.77)
	No	9	0.25 (0.11; 0.39)
	Unclear	5	0.14 (0.03; 0.25)
Comparator exposure arm			
Alignment follow-up – eligibility	Yes	21	0.58 (0.42; 0.74)
	No	11	0.31 (0.16; 0.46)
	Unclear	4	0.11 (0.01; 0.21)
Alignment follow-up – treatment	Yes	25	0.69 (0.54; 0.84)
	No	5	0.14 (0.03; 0.25)
	Unclear	6	0.17 (0.04; 0.29)
Alignment eligibility – treatment	Yes	20	0.56 (0.39; 0.72)
	No	12	0.33 (0.18; 0.49)
	Unclear	4	0.11 (0.01; 0.21)

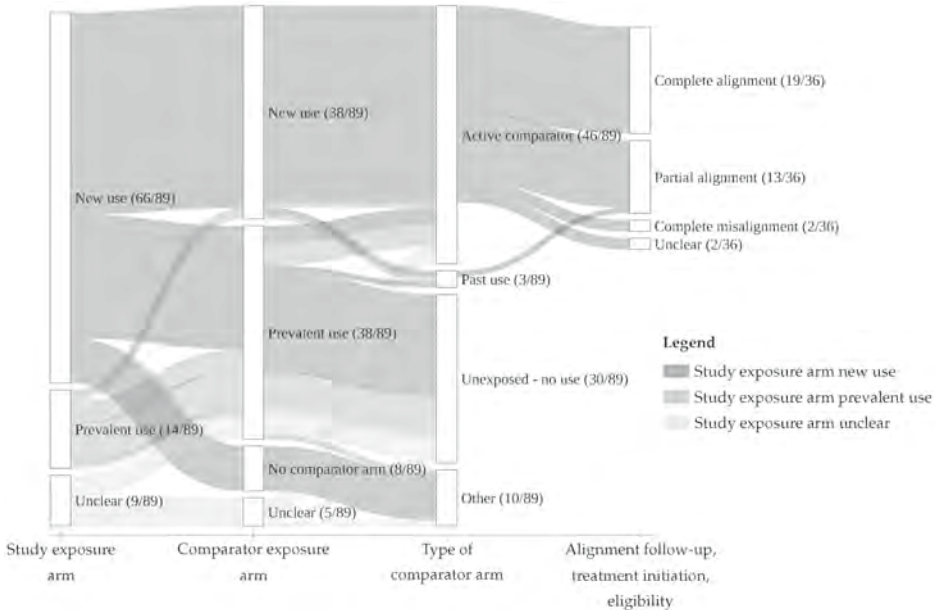


Figure 4. Frequency of reporting of implementation of new-user and prevalent-user design and type of comparator across the 89 included studies. For studies that reported implementing a new-user design, alignment of eligibility, treatment initiation and follow-up was scored 'completely aligned' when all three elements were reported to be aligned in both the active and comparator exposure arm; 'completely misaligned' when none of the elements were reported to be aligned in both the active and comparator exposure arm; 'unclear' when all three elements were unclear in both the active and comparator exposure arm; 'partial alignment' otherwise.

4.3 | Examples of good practice

Using examples from the 89 included studies, the next section illustrates how study designs that deviate from an archetypical pharmacoepidemiologic active-comparator new-user design could still provide estimates of the target treatment effect with a meaningful interpretation. We did not find any examples with a meaningfully defined study time origin among studies that contained a prevalent-user active-comparator arm.

Study design with non-user comparator arm

Korol and colleagues investigated whether initiation of spironolactone affected the risk of new onset diabetes in older patients with heart failure compared to not initiating spironolactone[34]. The patient cohort was defined by day of discharge of the first hospitalization for heart failure. The follow-up was started at the date of first dispensed prescription of spironolactone for the study exposure arm. The start of follow-up for unexposed comparator patients was inherited from the cohort entry date of the comparator and set to the time since hospital discharge from their matched comparator to establish a meaningful study time origin for non-users, given additional implementations to meet assumptions such as measuring sufficient confounders to invoke the exchangeability assumption (Table 4). Note that when an event-based cohort is established, resetting the start of follow-up at the moment of treatment initiation or comparable duration since diagnosis is essential to prevent introduction of immortal time bias[11].

Study design that anticipated immortal time

Chaignot and colleagues studied whether initiation of baclofen affected the risk of hospitalization and death compared to initiation of acamprosate in adults with an alcohol use disorder without comorbidities[35]. The patient cohort was defined by initiation of baclofen/acamprosate. To be eligible, patients had to receive at least two reimbursements for the same drug within 60 days after the first reimbursement, meaning that for included individuals, hospitalization/death could not have occurred before the second reimbursement was received. The start of follow-up was reset after the second prescription to prevent immortal time bias (Table 4). Note that the target causal effect changes by resetting start of follow-up. The study aims to identify the causal effect of baclofen compared to acamprostate given that everyone filled at least 2 prescriptions within 60 days and death was prevented during the time until they filled

a 2nd prescription. This interpretation is arguably more difficult to translate to clinical practice than a causal effect of initiating baclofen versus initiating acamprostate.

Study design that addressed time lags in start of follow-up

Belleudi and colleagues investigated whether switching from epoetin alpha (ESA α) to any other epoetin, compared to not switching, affected the risk of a blood transfusion or developing anaemia in chronic kidney disease patients[36]. The patient cohort was defined by initiation of ESA α. The follow-up was started at date of switching for the study exposure arm. A matched cohort was created to compare the risk of study outcomes in switchers versus non-switchers. The start of follow-up for non-switchers was matched to duration of ESA α treatment (± 30 days), thereby preventing time-lag bias (Table 4).

5 | Discussion

In our review of 89 pharmacoepidemiologic cohort studies of drug-outcome associations, 40% reported implementing a new-user design for both the study exposure arm and the comparator exposure arm, while 13% reported implementing a prevalent-user design in both arms, and 3 studies provided a rationale for including prevalent users. In studies that reported implementing a new-user design, we found there is room for improving alignment of meeting eligibility, treatment initiation, and start of follow-up, and reporting thereof.

It is not straightforward to understand the implications of misalignment of eligibility, treatment initiation, and start of follow-up in studies implementing a new-user design. Misalignment in the operationalization of the time origin in a study design can introduce immortal time bias or time-lag bias[3,5-7,11,12,17-19], but analytic methods can also help prevent these biases (e.g., analyzing treatment as a time-dependent variable as proposed by Suissa and Azoulay[13]). The validity of the chosen design and analysis is ideally assessed relative to the target causal effect. Since target causal effects were not often explicitly reported, we were not able to further assess implications of misalignment in the study time origin. It might have been possible to derive the target causal effect from information in the methods section in some studies. However, this would not contribute to assessment of the validity of the chosen design and analysis since target and operationalization would then overlap completely because of the reflexive definition of the target. When a target causal effect is not reported explicitly,

Table 4 Examples of design solutions for study time origin.

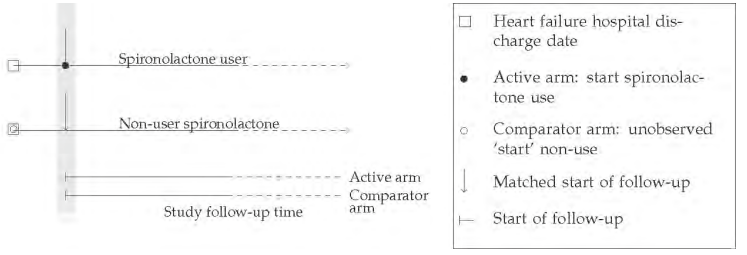
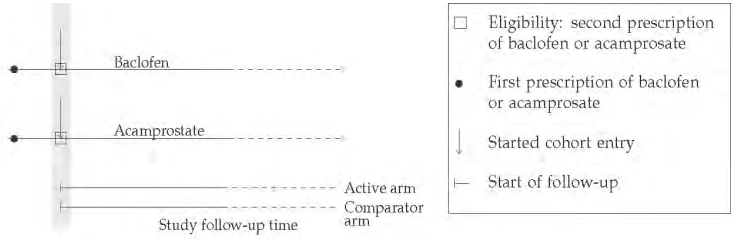
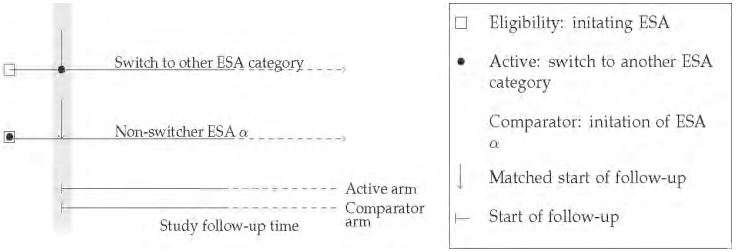
Research question	Designed time origin
Does initiation of spironolactone affect the risk of new-onset diabetes in older patients with heart failure compared to non-use of spironolactone?[34]	The patient cohort was defined by day of discharge of the first hospitalization for heart failure. For the study exposure arm, the follow-up was started at the date of first out-of-hospital dispensed prescription of spironolactone. The date of start of follow-up for unexposed comparator patients was matched to that of exposed patients on the time since hospital discharge axis to establish a meaningful study time origin for non-users. The authors did not report whether the non-user cohort was defined based on current exposure information or on future exposure information, i.e., whether non-users could still start spironolactone after their inherited date of start of follow-up or had to be unexposed during the entire study follow-up. The latter could result in a comparator cohort that is restricted to individuals who never had an indication for the treatment, which does not necessarily match the causal contrast of interest[37]. 

Table 4. *(continued)*

Research question	Designed time origin	
Does initiation of baclofen affect the risk of hospitalization and death compared to initiation of acamprosate in adults with an alcohol use disorder without comorbidities?[35]	The patient cohort was defined by initiation of baclofen/acamprosate. To be eligible, patients had to have received at least a second reimbursement for the same drug within 60 days after the first reimbursement. The start of follow-up was reset after the second prescription to prevent immortal time bias. The study thus estimates the causal effect of baclofen compared to acamprostate given that everyone filled at least 2 prescriptions within 60 days and death was prevented in the time until they filled a 2nd prescription.	
Does switching from epoetin alpha (ESA α) to any other epoetin, compared to not switching, affect the risk of a blood transfusion or developing anaemia in chronic kidney disease patients?[36]	The patient cohort was defined by initiation of ESA α. The follow-up was started at date of switching for the study exposure arm. A matched cohort was created to compare the risk of study outcomes in switchers versus non-switchers. The start of follow-up for non-switchers was matched to duration of ESA α treatment (± 30 days), thereby preventing time-lag bias (in matching, other covariates were considered as well).	

it is unclear which treatment effect the study aims to estimate, making it impossible to assess the impact of misalignment of eligibility, treatment initiation, and start of follow-up on validity of the study based on what is reported in the article. On the other hand, providing a concise and explicit definition of a target causal effect is a challenging task.

Our findings are in line with previous studies that investigated the implementation of the new-user design in specific patient domains. Yoshida and colleagues reviewed cohort studies investigating the association between use of disease-modifying antirheumatic drugs and either risk of infections (52 studies) or risk of cancers (15 studies) published between 2005 - 2015[15]. Forty percent of the studies on infection risk and 27% of the studies on cancer risk implemented a new-user active- comparator design, which is similar and lower, respectively, compared to the proportions found in our study, which covered a wider range of research areas. Suissa and Azoulay presented examples of observational studies investigating the association between metformin and cancer that suffered from immortal time bias, time-lag bias, or time-window bias[13]. Time-window bias can be an issue in case-control analysis and was not addressed here, because we only included cohort studies.

Based on our observations, it is our view that choosing a meaningful time origin is a more fundamental component of the study design than the distinction between new or prevalent users alone. Even when a new-user design was implemented, some of the articles we reviewed defined the study origin ambiguously. Reporting guidelines, such as RECORD-PE[38], state that study entry criteria and the order in which these criteria were applied to identify the study population should be clearly described. Indicating that a new-user design was implemented is insufficient to justify validity of a study design and time origin.

Our study had limitations. We focused on study-design approaches to define a meaningful study time origin. Although data analysis approaches can establish correct allocation of follow-up time as well[29,39], we did not assess them in our review. Misalignment of eligibility, treatment initiation, and start of follow-up may be appropriate when exposures are evaluated in a time-dependent manner. Four of the studies that reported implementing a new-user design studied a time-dependent exposure, thereby possibly adjusting for any misalignment in the study design. In our review, we assessed how frequently new-user and prevalent-user designs were implemented based on the reporting in original articles. It was not always possible to distinguish between lack of reporting and lack of implementation. Our results should therefore be interpreted as a summary of reporting practices on study time origin in

six journals. A final limitation is that our search was restricted to a convenience sample of six journals. Arguably, the six selected journals are representing the higher impact, specialist pharmacoepidemiology journals and results may therefore overestimate the quality of reporting of pharmacoepidemiologic studies in general.

The following recommendations for the design of pharmacoepidemiologic studies follow from our work. Reporting the motivation for a chosen study design and providing information on the extent to which moment of meeting eligibility criteria, treatment initiation, and start of follow-up are aligned improves the transparency and validity of research. We re-emphasize the importance of the recommendation by Schneeweiss and colleagues[40] to provide a design diagram, depicting a study's key temporal anchors and their relation to each other. When the target causal effect is unknown, it is difficult to assess whether study design and analysis are suitable for providing a meaningful estimate of the treatment effect of interest, in particular for time-dependent exposures. We recommend to explicitly report the causal contrast that is targeted in a separate statement at the beginning of the methods section. The definition of the target causal effect ideally concisely states the target population, the treatment strategies that are compared and how they are contrasted, and the outcome assessment (what and when). The causal contrast then explicates which effect is of interest (for example, an intention-to-treat effect, a per-protocol effect, an effect of treatment duration, or a comparison of treatment regimens)[26]. It should be unambiguous from this statement which future treatment decision can be informed by the study findings. Only when this information is clearly reported, the agreement can be assessed between target causal effect and applied study design and data analysis.

References

1. Johnson ES, Bartman BA, Briesacher BA, et al. The incident user design in comparative effectiveness research. Effective Health Care Program Research Report No. 32. (Prepared under Contract No. HHSA290200500161). AHRQ Publication No. 11(12)-EHC054-EF. *Rockville, MD: Agency for Healthcare Research and Quality.* 2012.
2. Yang W, Zilov A, Soewondo P, Bech OM, Sekkal F, Home PD. Observational studies: going beyond the boundaries of randomized controlled trials. *Diabetes Research and Clinical Practice.* 2010;88:S3-S9.
3. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Current Epidemiology Reports.* 2015;2(4):221-228.
4. Wang SV, Schneeweiss S, Berger ML, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1. 0. *Value in Health.* 2017;20(8):1009-1022.
5. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *American Journal of Epidemiology.* 2003;158(9):915-920.
6. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of Clinical Epidemiology.* 2016;79:70-75.
7. Johnson ES, Bartman BA, Briesacher BA, et al. The incident user design in comparative effectiveness research. *Pharmacoepidemiology and Drug Safety.* 2013;22(1):1-6.
8. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value in Health.* 2009;12(8):1053-1061.
9. Roberts AW, Dusetzina SB, Farley JF. Revisiting the washout period in the incident user study design: why 6–12 months may not be sufficient. *Journal of Comparative Effectiveness Research.* 2015;4(1):27-35.
10. Vandenbroucke J, Pearce N. Point: incident exposures, prevalent exposures, and causal inference: does limiting studies to persons who are followed from first exposure onward damage epidemiology? *American Journal of Epidemiology.* 2015;182(10):826-833.
11. Suissa S. Immortal time bias in pharmacoepidemiology. *American Journal of Epidemiology.* 2008;167(4):492-499.
12. Platt R, Hutcheon J, Suissa S. Immortal time bias in epidemiology. *Current Epidemiology Reports.* 2019;6(1):23-27.
13. Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care.* 2012;35(12):2665-2673.
14. Hempenius M, Luijken K, de Boer A, Klungel O, Groenwold R, Gardarsdottir H. Quality of reporting of drug exposure in pharmacoepidemiological studies. *Pharmacoepidemiology and Drug Safety.* 2020;29(9):1141-1150.
15. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nature Reviews Rheumatology.* 2015;11(7):437-441.
16. Perrio M, Waller PC, Shakir SA. An analysis of the exclusion criteria used in observational pharmacoepidemiological studies. *Pharmacoepidemiology and Drug Safety.* 2007;16(3):329-336.
17. 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. *International Journal of Epidemiology.* 2020;49(5):1719-1729.
18. Edwards JK, Hester LL, Gokhale M, Lesko CR. Methodologic issues when estimating risks in pharmacoepidemiology. *Current Epidemiology Reports.* 2016;3(4):285-296.
19. Farewell V, Cox D. A note on multiple time scales in life testing. *Journal of the Royal Statistical Society: Series C (Applied Statistics).* 1979;28(1):73-75.

20. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. *Developing a protocol for observational comparative effectiveness research: a user's guide*. 2013.
21. Hernán MA. Counterpoint: epidemiology to guide decision-making: moving away from practice-free research. *American Journal of Epidemiology*. 2015;182(10):834-839.
22. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *Journal of Clinical Epidemiology*. 2010;63(1):64-74.
23. Luijken K, Spekreijse JJ, van Smeden M, Gardarsdottir H, Groenwold RHH. The use of incident and prevalent-user designs in pharmacoepidemiology: a systematic review of the literature. 2020. osf.io/wn5ad.
24. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*. 2009;6(7):e1000097.
25. Moser CA. Quota sampling. *Journal of the Royal Statistical Society Series A (General)*. 1952;115(3):411-423.
26. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*. 2016;183(8):758-764.
27. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *British Medical Journal*. 2016;355.
28. Brookhart MA. Counterpoint: the treatment decision design. *American Journal of Epidemiology*. 2015;182(10):840-845.
29. Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiology and Drug Safety*. 2017;26(4):459-468.
30. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960;20(1):37-46.
31. Campbell NL, Lane KA, Gao S, Boustani MA, Unverzagt F. Anticholinergics influence transition from normal cognition to mild cognitive impairment in older adults in primary care. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*. 2018;38(5):511-519.
32. Harding BN, Weiss NS, Walker RL, Larson EB, Dublin S. Proton pump inhibitor use and the risk of fractures among an older adult cohort. *Pharmacoepidemiology and Drug Safety*. 2018;27(6):596-603.
33. Young JC, Lund JL, Dasgupta N, Jonsson Funk M. Opioid tolerance and clinically recognized opioid poisoning among patients prescribed extended-release long-acting opioids. *Pharmacoepidemiology and Drug Safety*. 2019;28(1):39-47.
34. 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? *European Journal of Clinical Pharmacology*. 2019;75(6):837-847.
35. Chaignot C, Zureik M, Rey G, Dray-Spira R, Coste J, Weill A. Risk of hospitalisation and death related to baclofen for alcohol use disorders: Comparison with nalmefene, acamprosate, and naltrexone in a cohort study of 165 334 patients between 2009 and 2015 in France. *Pharmacoepidemiology and Drug Safety*. 2018;27(11):1239-1248.
36. Belleudi V, Trotta F, Addis A, et al. Effectiveness and safety of switching originator and biosimilar epoetins in patients with chronic kidney disease in a large-scale Italian cohort study. *Drug Safety*. 2019;42(12):1437-1447.
37. Lund JL, Horváth-Puhó E, Szépligeti SK, et al. Conditioning on future exposure to define study cohorts can induce bias: the case of low-dose acetylsalicylic acid and risk of major bleeding. *Clinical Epidemiology*. 2017;9:611.
38. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *British Medical Journal*. 2018;363.
39. Rachet B, Abrahamowicz M, Sasco A, Siemiatycki J. Estimating the distribution of lag in the effect of short-term exposures and interventions: adaptation of a non-parametric regression spline model. *Statistics in Medicine*. 2003;22(14):2335-2363.
40. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical depiction of longitudinal study designs in health care databases. *Annals of Internal Medicine*. 2019;170(6):398-406.