

Incidence in pharmacoepidemiology—Basic definitions and types of misclassification

Mikael Hoffmann¹  | Henrik Støvring² 

¹Health Care Analysis, Division of Society and Health, Linköping University, Sweden & The NEPI Foundation, Stockholm, Sweden

²Clinical Pharmacology, Pharmacy and Environmental Medicine, University of Southern Denmark, Odense, Denmark

Correspondence

Mikael Hoffmann, Health Care Analysis, Division of Society and Health, Linköping University, Sweden & The NEPI foundation, Stockholm, Sweden.
Email: mikael.hoffmann@liu.se

Funding information

NEPI foundation, Grant/Award Number: SE802400-2589

Abstract

The definition of a new case is a vital step in incidence studies in both epidemiology and pharmacoepidemiology, although with significant differences in methodology between the fields. We define and apply a framework for two different types of new cases of drug use, first-ever and recurrent, and show how the associated misclassifications related to length of run-in period can be expressed by the positive predictive value (PPV).

In the study, we consider individual-level dispensations of statins 2006–2019 for 1,017,058 individuals with at least one dispensation in 2019 in Sweden. The incidence proportion for statins for both sexes of all ages in Sweden 2019 varied from 17.4/1000 with a run-in of 8 months, 9.45/1000 with 5 years and 8.4/1000 with 10 years. The PPV was 49% with 8 months and 89% for 5 years using 10 years as gold standard.

We conclude that the interpretation of incidence and thus the selection of an appropriate run-in period, in pharmacoepidemiology, depends on whether first-ever use, recurrent treatment or both together (new cases) is the focus of the research question studied. At least five different misclassifications can be introduced depending on how incidence is defined.

KEYWORDS

incidence, misclassification, pharmacoepidemiology, run-in, statins

1 | BACKGROUND

The definition of a new case is a vital step in both epidemiology and pharmacoepidemiology, but significant differences in methodology exist between the fields. The aim of this article is to suggest a theoretical framework for describing different types of new cases of drug use, and related misclassifications, as a basis for more detailed definitions of incidence in pharmacoepidemiology. In

addition, we exemplify this with statins as an example of a treatment of a chronic condition.

1.1 | Incidence in epidemiology

In epidemiology, measures of disease frequency such as incidence and prevalence are well defined.¹ In the simplest case, these measures refer to a disease leading to immunity after the disease period (e.g., measles) or that

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Basic & Clinical Pharmacology & Toxicology* published by John Wiley & Sons Ltd on behalf of Nordic Association for the Publication of BCPT (former Nordic Pharmacological Society).

is chronic (e.g., diabetes mellitus type 1). Then, once a person has been a new case, an incident patient, he/she will not be able to acquire the disease again. This situation may be described by the so-called illness-death model (also known as the disability model).² In this model, individuals can be in one of three states: healthy (not yet diseased), diseased (immune) or deceased. In the simplest variant, a diseased individual cannot return to a healthy state and again be at risk of contracting the disease (see Figure 1).

Diseases with potential recovery require an expansion of the illness-recovery-death model. Recovered individuals can either be susceptible to falling ill again or be protected through immunization, natural immunity or vaccination. The protection can diminish over time, thus making the individual susceptible to the disease again.

In epidemiology, the incidence may be reported as follows:

- number of events,
- *incidence rate* (or person-time rate)—a rate of cases relative to the aggregate cumulative study time contributed by studied individuals, or
- *incidence proportion* (or cumulative incidence)—a proportion of new cases among the entire population at risk accrued over a time interval.³

The *incidence rate* is the ratio between the number of new cases of the disease during a specified time interval and the sum of all person-time at risk. The *incidence proportion* is the number of individuals in an initially disease-free population that becomes cases during a specified period divided by the total number of individuals in the population at the start of the studied period.

According to the definitions for both incidence rate and proportion, the denominator should be limited to the population at actual risk of developing the disease (i.e., those who are exposed to risk).^{1,4,5} In order to strictly adhere to the definition of disease incidence, it is thus necessary to exclude patients with the pre-existing disease through a look-back period. The length of this period will introduce a relative misclassification with an

overestimation of incidence rates, higher for a shorter look-back period.

The illness-death model would include healthy individuals who have not yet been diseased for the first and only time in their life. In the illness-recovered-death model, the population at risk would include those not yet afflicted by disease and those recovered who are susceptible to the same disease once again.

Identifying the actual individuals who have recovered is often complicated, especially regarding excluding the period in which they are immune. Thus, it is not uncommon to use the term “*incidence*” to refer neither to the incidence rate nor the incidence proportion but to a ratio between new cases and the total population, including those not at risk. The result will be an apparently lower incidence rate or incidence proportion due to the inflated denominator.^{6,7}

1.2 | Incidence in pharmacoepidemiology

The identification of new and current users of a drug is essential in pharmacoepidemiology.

The original simple model based on infections need to be extended in order to be applicable to the dispensation of drugs as a proxy for drug treatment and being treated or not treated with a drug instead of having an acute or a chronic condition. This article presents an adaptation of the basic model for incidence for the purpose of calculating incidence in pharmacoepidemiology.

Data about dispensations linked to unique individuals have, over the last few decades, led to the rapid development of both drug utilisation studies and more analytic pharmacoepidemiology to evaluate the effects and side effects of pharmaceutical treatment.⁸ A pivotal question is how to model treatment over time from data about dispensed drugs, and in extension, how to estimate point prevalence, initiation and duration of treatment with varying methodologies depending on the research question and healthcare setting.⁹

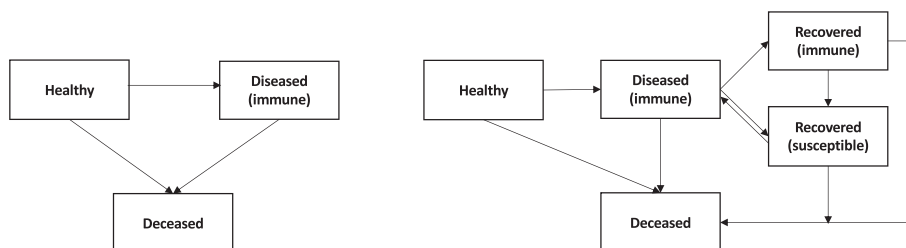


FIGURE 1 Simple illness-death model (left) compared with a suggested expanded illness-recovery-death model (right)

Hospital records with ordered and administered doses documented over time allow identification of initiation of drug treatment, change of doses, and stop of treatment. This is also possible in prospective studies documenting actual doses administered by the patient at home. However, studies of drug utilisation are usually register-based studies of dispensed prescriptions at pharmacies. This introduces several methodological issues. The actual initiation of the drug therapy might have started later than the actual dispensation. The drug might not be taken at all. Dose changes might occur without new dispensations. The time the dispensed amount will cover treatment is not always defined. Previous hoarding, as well as partial non-compliance, might extend treatment time. Finally, termination of treatment frequently occurs before the possible duration of treatment based on dose and dispensed amount.

In drug utilisation studies, the ability to identify new users from large databases of dispensed pharmaceuticals is of particular interest as it allows analysis of early changes in prescribing. The first dispensation of the first-time prescription to a patient is of particular interest since it is a direct consequence of the physician's operational decision about diagnosis and treatment. The first dispensation of a renewed prescription is also of more interest than repeated dispensations of the same prescription since each renewed prescription represents a new decision by the physician. Another aspect of identifying a new user is being able to calculate the cumulative amount of exposure when this can be linked to the pharmacological effect of a drug. Market trends in total sales volume or the number of treated patients are consequences of both new users and treatment duration. Identifying new users and individuals stopping treatment can thus provide additional information that might enhance trend analysis in drug utilisation studies.

A *run-in period* (sometimes also called *washout period*) is commonly used to differentiate between a dispensation indicating a *new case* of drug use and one representing a continuation of treatment.

A commonly chosen run-in period length is 1 year (365 days) for different drugs such as antipsychotics,^{10,11} psychotropics,¹² benzodiazepines¹³ and direct oral anticoagulants.¹⁴ For drugs with an evident seasonal variation, a run-in of 1 year may avoid unnecessary bias. However, both shorter^{15,16} and more extended run-in periods are also used.¹⁷ For proton pump inhibitors, up to 24 months run-in¹⁸ and for opioids up to 3–5 years^{19,20} have been used. In many studies, there is no expressed reason for, or validation of, the chosen run-in period, although there are a few exceptions, such as the choice of 3 years

motivated by the need to identify “*apparent new users of prescription opioids*.”¹⁹

As it is difficult to define an appropriate length of the run-in period, even with good clinical insight into the specific setting, the parametric waiting time distribution (WTD) has been suggested as an analytical approach to estimate the duration of dispensing.²¹ Its simplest form is solely based on observed dates of dispensations. It then provides an estimate of the dispensing duration defined as the time within which a specified percentage of all users in ongoing treatment will have returned to collect their next dispensation. In addition, the method provides an estimate of the proportion of *new cases* of drug treatment among all with an observed dispensation during the observation period, although it does not differentiate between *first-ever use* and *recurrent treatment*. However, the WTD could also be used to inform the choice of a run-in period, as this may be chosen to exceed, for example, the 99th percentile of the estimated duration of dispensings. This would imply that less than 1% of patients in ongoing treatment are classified as *new cases* of drug treatment based on such a run-in period.

The National Board of Health and Welfare provides annual incidence proportions for some substance groups as publicly available drug consumption statistics. The methodology is an early variant of WTD calculated as the average monthly number of individuals during the last 4 months of a calendar year that have not filled a prescription during the earlier months of the calendar year. It can be crudely interpreted as a run-in time of between 10 and 11 months.²² The methodology does not consider a seasonal variation of disease or healthcare interventions. In general, fewer patients with non-acute conditions will be seen by a physician during the summer months, and thus, fewer treatments for chronic conditions will be initiated. Vacation plans can sometimes lead to patients filling their prescriptions earlier than otherwise. Thus, the methodology probably overestimates the incidence proportion of, for instance, statins (C10AA HMG CoA reductase inhibitors and fixed combinations of HMG CoA reductase inhibitors in C10BA). The reported incidence proportion of statins for all ages in 2019 according to the National Board of Health and Welfare was 15.8/1000 women and 19.5/1000 men.²³

Estimation of the WTD proceeds from discarding all dispensing of individual patients after their first dispensation within an interval (ordinary WTD), or of all dispensations preceding their last within the interval (reverse WTD). The length of the interval is routinely taken to be 1 year in most instances, as the method is robust as to regards varying the length, as long it is not very short (less than 6 to 8 months, say). Based on

renewal process theory, it has been shown that a parametric likelihood analysis can be applied to estimate percentiles of the so-called interarrival distribution for patients in ongoing treatment, when estimating prescription durations.²¹ In addition, the procedure provides a direct estimate of the proportion of users, which have initiated treatment during the observation period among all with a dispensing within the period. The estimation procedure has been implemented in Stata in a freely available software package (type—search wtdttt—in Stata to locate and download).²⁴ In an extension, the method allows the inclusion of characteristics of the dispensed medication (amount, type, etc.) and the patient (age, sex, etc.) as predictors for dispensing durations. However, when used to inform the length of the run-in period, a single estimate of the maximal dispensing duration may be preferable.

The main advantage of using the WTD to estimate dispensing duration, and in turn, the length of a run-in period to identify *new cases* of drug treatment, is that its estimate is based on the actual usage pattern of the population considered as studied through dispensing dates. The main limitation of the procedure is that it, by definition, cannot establish a maximal dispensing duration within which 100% of all patients in ongoing treatment will have returned to collect their next dispensation—such a limit is formally defined to be at infinity. This is, however, likely to be of minor practical importance, as a 99th percentile, or 99.5, will suffice for most settings as long as *new cases* of drug treatment are studied and not *first-ever use*. More work needs to be done to establish how the

choice of percentile affects misclassification rates of the corresponding run-in period.

2 | METHODOLOGICAL CONSIDERATIONS WHEN DEFINING A NEW CASE OF DRUG USE

2.1 | Differentiating between first-ever use and recurrent treatment

When studying episodes of drug use, or dispensations as proxies for drug use, then extending the illness-death model to consider different states and transitions between states is crucial (see Figure 2). In this extended model, an untreated patient can be either a *never user* (A) or an untreated *previous user* (C). A treated patient can be either a *first-ever user* or a *recurrent user*.

The dispensation of the studied drug to a *never user* (A) will then be the first dispensation during the patient's life, that is, identifying the transition $A \rightarrow B$ of the individual to the state of *first-ever user* (B). The same individual will remain in the *first-ever user* (B) state until the treatment is stopped and the individual becomes a *previous user* (C) with the transition $B \rightarrow C$.

Correspondingly, the dispensation of the studied drug to a *previous user* (C) will define the transition $C \rightarrow D$ of the individual to the state of *recurrent user* (D). The individual will remain in this state until the treatment ends, implying a return to the state of the *previous user* (C).

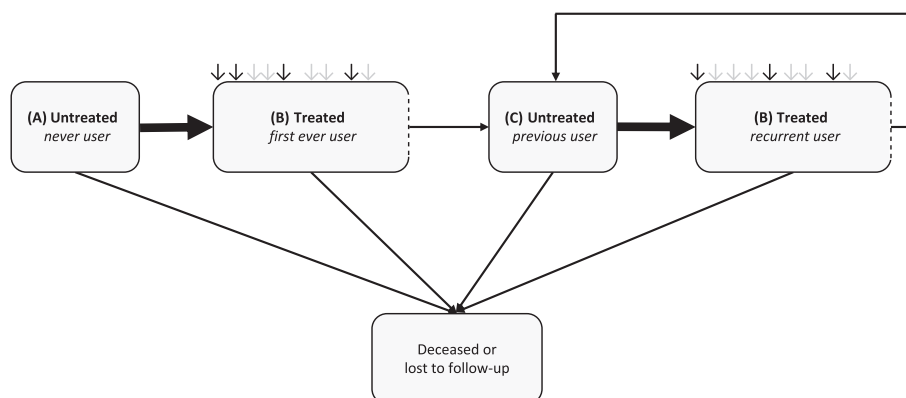


FIGURE 2 Model for repeat treatment in pharmacoepidemiology. States refer to actual use during the defined period. Transitions between states refer to different situations. In pharmacoepidemiology, transitions from $A \rightarrow B$ and $C \rightarrow D$ (bold arrows) define new cases. $A \rightarrow B$ represent first-ever use, while $C \rightarrow D$ represent recurrent treatment. The dotted border depicts uncertainty about when the treatment episode ends when using dispensations as proxies for drug treatment. Vertical arrows represent dispensations (black = first dispensation of a prescription, grey = repeat dispensations of the previous prescription) in hypothetical first-ever and recurrent user episodes

When dispensations are used as proxies for drug treatment, the transitions $B \rightarrow C$ and $D \rightarrow C$ are not well defined, as indicated by the dotted line in Figure 2. The treatment can end before the dispensed amount is administered, and the duration could be extended due to past hoarding, a reduced dose or partial non-compliance. Thus, in pharmacoepidemiology, the time the dispensation results in administration will need to be estimated. This can be done by analysing the prescribed dose and amount dispensed or based on the average interval between dispensations for the current medication in the studied setting. This time, the *assumed duration* of treatment after a dispensation will have to be added to the date of the last dispensation during B or D to define the beginning of C. This introduces an intermediate state of uncertainty about whether the patient is under treatment before becoming a previous user. A decision has also to be made on how long the actual interval without treatment has to be from a clinical standpoint for a new dispensation to be interpreted as a *recurrent treatment* and not a continuation of the treatment. This time should be added to the *assumed duration* to define a run-in period for identifying a clinically relevant *new case* of *recurrent treatment*.

In addition to the assumed duration of the treatment after a dispensation in each interval, there are other factors to consider. A patient can be considered to be under treatment with a drug under an extended period after the last administered dose, for instance, for a drug such as fluoxetine with a long half-life for the substance and its active metabolites.²⁵ From a clinical perspective, a new treatment episode can also be defined as starting only after an extended time with no treatment.

2.2 | Different types of misclassifications depending on the run-in

The length of the run-in can introduce different types of misclassifications of *new cases*, *first-ever use* and *recurrent treatment*.

2.2.1 | Misclassification due to immigration

Defining a *new case* by the lack of recorded dispensations during a defined run-in period is dependent on identifying all dispensations during run-in for the studied individuals. This is not possible for individuals entering the study population by immigration when studying incidence proportion in a general population register without specific record-linkage. For these individuals, the first

recorded dispensation after immigrating will be identified as a *new case* of drug use whether or not this is correct. For incidence rate, migration is handled through the study of actual person-time.²⁶

Censoring all individuals who immigrated during the run-in before an index dispensation will avoid this misclassification. Instead, it will introduce an underestimation of the true incidence of new cases if immigrants are not censored from the denominator. However, since the number of immigrants in most cases is only a fraction of the population and only a fraction of the immigrants are prevalent users, the misclassification in most cases is low.

2.2.2 | Continued treatment misclassified as recurrent treatment

If the run-in is close to the duration of treatment after a dispensation, that is, if the actual duration of treatment after a dispensation and the clinically relevant period of non-treatment needed to identify a dispensation as a new treatment episode added together is short, then a repeat dispensation might be misclassified as a *recurrent treatment*.

2.2.3 | Recurrent treatment misclassified as continued treatment

A *recurrent treatment* might be misclassified as continued treatment if the run-in is longer than the duration of treatment after a dispensation plus the time needed without treatment to define a *new case* of drug treatment. In addition, during the run-in, a patient cannot be identified as a *new case* of drug treatment whether or not this is true in the individual case.

When drug treatment is halted (either as planned, as a separate decision made by the prescriber or through lack of compliance), the individual will transition into *previous use* ($B \rightarrow C$ or $D \rightarrow C$) immediately in real life. Nevertheless, when using dispensations as proxies for the individual's treatment status, the exact point in time where this transition will occur is not observed when only the actual dispensed amount and the prescribed dose is known. A patient who stops taking the drug shortly after a dispensation will thus not qualify as a *new case* for the remainder of the run-in period.

The degree of misclassification in I and II cannot easily be calculated, since in each case the duration of treatment has to be estimated because it cannot be identified through information about dispensations only.

2.2.4 | First ever use misclassified as recurrence of treatment

The appropriate run-in for studying new cases should consider both the *assumed duration* of the last dispensation and the clinically relevant period of non-treatment needed to classify a new episode of drug treatment as a *recurrent treatment*. This run-in will identify new cases but not separate recurrent treatment from first-ever use. In order to identify only cases of recurrent treatment, the number of first-ever use has to be subtracted from new cases.

2.2.5 | Recurrent treatment misclassified as first-ever use

To define an actual *first-ever use*, a run-in period would need to cover a life-long observation period or the period since the drug was introduced into the market being studied.

The longer the run-in period is, the higher the ratio of *first-ever use* to *recurrent treatment*. Thus, the run-in period has to be chosen based not only on the characteristics of the drug studied but also on the clinical research question, that is, whether the focus is on

- first-ever use
- recurrent treatment
- or all *new cases* of drug treatment episodes, that is, both *first-ever use* and *recurrent treatment*

Any chosen run-in shorter than the individual's life or period since drug market entrance will introduce misclassification where some *new cases* will be instances of *recurrent treatment* instead of *first-ever use*.

The degree of misclassification depends on a range of factors such as drug type, gender, age group, study period, and the historical prescriptions and dispensations of the drug, which can be quantified in different ways.^{27–29} For instance, a woman in her sixties will have a higher probability to have been treated with menopausal hormone therapy during the last 10 years compared with a woman in her forties. This introduces a possible bias with a lower positive predictive value given the same run-in for *first-ever use* among elderly and a corresponding higher relative misclassification among older women.³⁰

The proportion of *new cases* of drug use, defined by a chosen run-in period, that represents *first-ever use* can be assessed by calculating the *positive predictive value*³¹ compared to an extended observation time, for instance, 10 years as a pragmatic gold standard. In reverse, a

desired positive predictive value can guide the selection of a suitable run-in if the goal is to identify the *first-ever use* of drugs.

New cases selected through a long run-in will have a high positive predictive value for identifying a *new case* of drug use as *first-ever use*. However, a long run-in will also misclassify some cases of *recurrent treatment* as not being a *new case* of drug use, see Misclassification III.

3 | EXAMPLE—INCIDENCE OF STATIN USE IN SWEDEN 2019

3.1 | Material and methods

To exemplify the concepts introduced and discussed above, we have studied the first dispensation during 2019 of any statin (C10AA HMG CoA reductase inhibitors and fixed combinations of HMG CoA reductase inhibitors in C10BA) to individuals in Sweden, and for each such dispensation all previous dispensation to the same individual since 1 July 2005. Statins were chosen as an example for demonstration purpose since this is a common and well-studied treatment of a chronic condition.

The Swedish Prescribed Drug Registry data were extracted as patient-level data, fully anonymised and classified as statistics by the National Board of Health and Welfare.³² Drugs were classified according to Anatomical Therapeutic Chemical (ATC) classification system in 2020.^{33,34}

All first individual occurrences of the dispensation of C10AA HMG CoA reductase inhibitors and fixed combinations of HMG CoA reductase inhibitors in C10BA during 2019 were extracted (*n* of individuals = 1 017 058) together with the ATC code and number of days since the last dispensation of the same ATC code. In addition, the number of days since the last dispensation of any other of the studied ATC codes was obtained together with information on gender, age (5-year intervals up to 85+), and Swedish citizen status on 1 January in 2009 and 2019. Stata²⁴ was used for all analyses of data including the waiting-time distribution.

Incidence proportion was calculated with the number of new cases (either first-ever or recurrent treatment) defined by different run-in periods as the nominator and the population in the beginning of the year as denominator. Positive predictive value was calculated as the percentage of the numbers of new cases defined by a chosen run-in as nominator and the number of new cases defined by a pragmatically selected run-in of 10 years.

Number of new cases were also calculated using a freely available software for waiting time distribution as described in Background.

TABLE 1 New cases for statins as a group, by WTD and different run-in periods

	WTD		8 months		12 months		16 months		3 years		5 years		10 years	
	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV
Both sexes <i>n</i> = 1,017,058	15.27 (15.15–15.39)	55%	17.35 (17.27–17.43)	49%	13.71 (13.64–13.78)	62%	12.24 (12.17–12.31)	69%	10.26 (10.20–10.33)	82%	9.45 (9.39–9.51)	89%	8.44 (8.38–8.49)	
Women <i>n</i> = 439,009	13.55 (13.39–13.71)	56%	15.67 (15.56–15.78)	49%	12.48 (12.38–12.57)	61%	11.17 (11.07–11.26)	68%	9.38 (9.29–9.46)	81%	8.60 (8.52–8.68)	88%	7.61 (7.53–7.68)	
Men <i>n</i> = 578,049	16.95 (16.77–17.13)	55%	19.02 (18.90–19.14)	49%	14.94 (14.83–15.04)	62%	13.30 (13.20–13.40)	70%	11.14 (11.05–11.23)	83%	10.29 (10.20–10.37)	90%	9.26 (9.18–9.34)	
Women 45–64 <i>n</i> = 99,568	19.21 (18.85–19.57)	49%	16.46 (16.26–16.66)	58%	13.74 (13.56–13.93)	69%	12.49 (12.31–12.66)	76%	10.81 (10.64–10.97)	88%	10.17 (10.01–10.32)	93%	9.49 (9.34–9.65)	
Men 45–64 <i>n</i> = 167,829	28.21 (27.77–28.65)	47%	24.09 (23.85–24.33)	55%	19.74 (19.52–19.96)	67%	17.82 (17.62–18.03)	74%	15.18 (14.99–15.36)	87%	14.24 (14.06–14.43)	93%	13.19 (13.02–13.37)	
Women 65+ <i>n</i> = 332,545	38.27 (37.65–38.89)	53%	46.89 (46.49–47.29)	43%	36.19 (35.84–36.54)	56%	32.02 (31.69–32.36)	63%	26.35 (26.05–26.65)	76%	23.69 (23.40–23.97)	85%	20.10 (19.84–20.36)	
Men 65+ <i>n</i> = 394,163	47.41 (46.68–48.14)	50%	55.88 (55.42–56.35)	42%	41.93 (41.52–42.33)	56%	37.77 (36.39–37.15)	62%	30.16 (29.81–30.50)	78%	27.29 (26.96–27.62)	86%	23.60 (23.29–23.90)	

Note: All ages, 45–64 years, 65+ years. Positive predictive value (PPV) compared with a run-in of 10 years. Not censored for immigration.

In addition, data on the incidence of C10AA HMG CoA reductase inhibitors and fixed combinations of HMG CoA reductase inhibitors in C10BA during 2019 (all ages and genders) were collected from the National Statistics of Drug Consumption at the website of the National Board of Health and Welfare. Incidence was calculated as the mean of new users during the last 4 months of the calendar year.

3.2 | Results

Depending on the length of the run-in period, the incidence proportion for statins as a group for both sexes of all ages in Sweden 2019 varied from 17.4/1000 with a run-in of 8 months, 9.45/1000 with a 5-year run-in and 8.4/1000 with a 10-year run-in (Table 1 and Figure 3). The corresponding positive predictive value, PPV, for identifying first-ever use (pragmatically defined as no dispensations during 10 years prior to the index date) were 49% with a run-in of 8 months and 89% with a run-in of 5 years. The trends were similar for the different age and sex groups.

When censoring immigrated individuals by only analysing individuals residing in Sweden in both 2009 and 2019, the incidence proportion for both sexes and all ages fell by 5.1% to 6.5% depending on the run-in (Tables 1 and 2 and the supporting information). The discrepancy varied between 1.8% (women 45–64 with 12 months run-in) and 9.0% (men 45–64 with 10 years run-in).

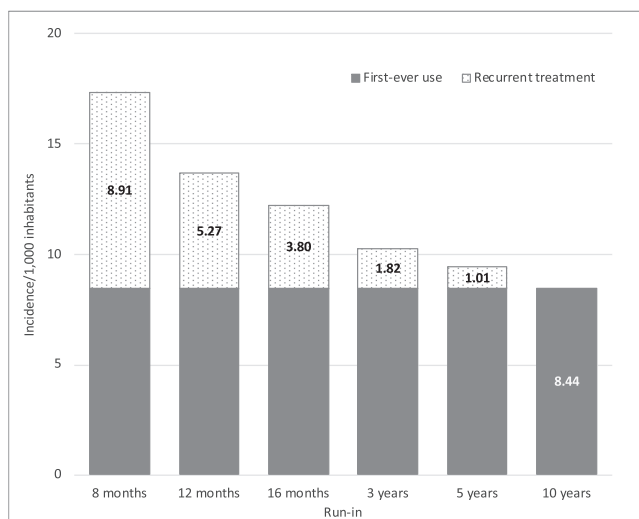


FIGURE 3 Incidence proportion of statin use as first-ever use and recurrent treatment in individuals per 1000 inhabitants 2019, all ages and both sexes, with different run-in. First-ever use pragmatically defined as no previous use during 10 years before the index date

The incidence proportion assessed with the waiting-time distribution function in STATA for both sexes, all ages, was 15.27/1000 patients (13.55/1000 women and 16.95/1000 men), an intermediate between 8 and 12 months of run-in.

4 | DISCUSSION

In pharmacoepidemiology, a *new case* of drug treatment can either represent a case of *first-ever use* or *recurrent treatment*. When a run-in period is used (either explicitly defined or implicitly used as in a waiting-time distribution), the results will be decided by the length of the run-in period. In this paper, we discuss five different types of misclassifications. The most crucial issue is to differentiate between *first-ever use* and *new cases* of drug treatment (or the sum of *first-ever use* and *recurrent treatment*). Thus, it is crucial to define misclassification due to the chosen run-in period compared with a gold standard when using incidence to describe *first-ever use*.

The positive predictive value for a specific length of a run-in period compared to a long run-in period approximating no earlier use is an easy-to-communicate description of the fraction of *first-ever use*. It quantifies the relative misclassification and the influence of the run-in period when incidence is used to assess *first-ever use*. However, it is also helpful if there is a need to focus only on *recurrent treatment*. For example, if *first-ever use* is pragmatically defined as no dispensation for the preceding 10 years with a 49% PPV for the incidence of 17.4/1000 at 8 months run-in, then the incidence with the same run-in period will be almost evenly divided between *first-ever use* (49%) and *recurrent treatment* (1-PPV, or 51%).

A waiting-time distribution based on the original definition (studying the frequency of individuals with the first dispensation during the last few months of 12 months) emulates a run-in of only 8–12 months. In most cases, *new cases* will then be dominated by *recurrent treatment* and not *first-ever use*. This is shared by the parametric approach, although it provides more flexibility in the estimation with introduction of explanatory variables and randomly selected index time points for each individual.^{35,36} However, it is essential to discuss the balance between *first-ever use* and *recurrent treatment* in both cases and relate this to the research question. The free software for the parametric waiting time distribution is a versatile tool to estimate the proportion of *new cases*, which will represent a mixture of *first-ever use* and *recurrent treatment*.

We recommend against using a simplified waiting-time distribution studying the mean numbers of new

TABLE 2 New cases for statins as a group, all ages, by WTD and different run-in periods. Censored for immigration

	8 months		12 months		16 months		3 years		5 years		10 years	
	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV	New cases/1000	PPV
Both sexes <i>n</i> = 992,022	16.47 (16.39–16.55)	48%	12.98 (12.91–13.05)	61%	11.57 (11.51–11.64)	68%	9.68 (9.62–9.74)	82%	8.88 (8.83–8.94)	89%	7.89 (7.83–7.94)	
Women <i>n</i> = 429,685	15.00 (14.90–15.11)	48%	11.92 (11.83–12.02)	60%	10.66 (10.57–10.75)	68%	8.94 (8.86–9.02)	81%	8.17 (8.10–8.25)	88%	7.20 (7.12–7.27)	
Men <i>n</i> = 562,337	17.92 (17.81–18.04)	48%	14.03 (13.92–14.13)	61%	12.47 (12.38–12.57)	69%	10.41 (10.33–10.50)	82%	9.59 (9.50–9.67)	89%	8.57 (8.49–8.65)	
Women 45–64 <i>n</i> = 94,910	15.28 (15.09–15.47)	57%	12.76 (12.59–12.94)	69%	11.59 (11.42–11.76)	76%	10.03 (9.88–10.19)	87%	9.42 (9.26–9.57)	93%	8.76 (8.61–8.90)	
Men 45–64 <i>n</i> = 159,516	22.19 (21.96–22.41)	54%	18.15 (17.94–18.36)	66%	16.38 (16.18–16.57)	73%	13.92 (13.74–14.1)	86%	13.03 (12.85–13.21)	92%	12.00 (11.83–12.17)	
Women 65+ <i>n</i> = 328,941	46.01 (45.62–46.41)	43%	35.51 (35.16–35.86)	55%	31.43 (31.10–31.75)	62%	25.84 (25.54–26.14)	76%	23.21 (22.93–23.50)	85%	19.64 (19.38–19.90)	
Men 65+ <i>n</i> = 389,389	54.54 (54.08–55.00)	42%	40.91 (40.51–41.31)	56%	35.86 (35.48–36.23)	64%	29.39 (29.05–29.73)	78%	26.58 (26.26–26.90)	86%	22.91 (22.60–23.21)	

Note: All ages, 45–64 years, 65+ years. Positive predictive value (PPV) compared with run-in 10 years.

users of the last 4 months for a calendar year. This method is susceptible to seasonal variation and overestimates the new cases for drugs that are more frequently initiated before or after the summer period. This is evident in our study where the reported incidence proportions from the National Board of Health and Welfare for statins exceed even the result using 8 months as a run-in.

We also recommend specifying whether the reported incidence proportion represents *new cases* or *first-ever use* and providing an analysis of the relationship between *first-ever use* and *recurrent treatment* among *new cases*.

The incidence, or the number of new cases, will vary with the chosen run-in period (which will define the relative proportion between *first-ever use* and *recurrent treatment*), sex and age. This variation is a consequence of incidence being defined by current prescribing practices and how the drug has been available/prescribed historically to different populations. Thus, the choice of run-in period should be carefully considered for different sub-populations, drug treatments, geographical areas, countries and periods.

Immigration during the run-in period, if not censored for, will introduce misclassification of treated individuals as new cases. This misclassification will vary depending on the age distribution of immigrants and the prevalence of drug treatment among immigrants.

5 | CONCLUSIONS

New cases or incidence of drug treatment represents the sum of *first-ever use* and *recurrent treatment*. The length of the selected run-in period will define the relationship between these two different situations and will introduce a misclassification that can be bidirectional.

The interpretation of incidence, and thus the selection of an appropriate run-in period, in pharmacoepidemiology, depends on whether *first-ever use*, *recurrent treatment* or both together (*new cases*) is the focus of the research question studied.

Different misclassifications can be introduced depending on whether first-ever use, recurrent treatment or new cases are studied. The positive predictive value is a valuable tool to describe the fraction of *new cases* defined by a given run-in that are instances of *first-ever use*, and thus also the corresponding fraction of *recurrent treatment* episodes. Conversely, the length of the run-in can be defined by the desired positive predictive value.

ACKNOWLEDGEMENTS

The study was fully funded by the NEPI foundation (Stiftelsen NEPI – nätverk för läkemedelsepidemiologi, org# SE802400-2589). The NEPI foundation is a non-

profit, tax-exempt foundation initiated by the Swedish parliament in 1993 with the aim to support pharmacoepidemiology, health economics and drug information.

CONFLICT OF INTEREST

Both authors report no conflicts of interest according to ICMJE Disclosure form.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author, MH, upon reasonable request.

ORCID

Mikael Hoffmann  <https://orcid.org/0000-0001-9420-3316>

Henrik Støvring  <https://orcid.org/0000-0002-5821-2351>

REFERENCES

- Porta M. (Ed). *Dictionary of Epidemiology*. 6th ed. Oxford University Press; 2016.
- Hougaard P. Multi-state models: a review. *Lifetime Data Anal.* 1999;5(3):239-264. doi:10.1023/A:1009672031531
- Principles of Epidemiology in Public Health Practice. An Introduction to Applied Epidemiology and Biostatistics*. 3rd ed. Atlanta, GA, USA: Centers for Disease Control and Prevention (CDC); 2012. <https://www.cdc.gov/csels/dsepd/ss1978/ss1978.pdf>
- Griffiths RI, O'Malley CD, Herbert RJ, Danese MD. Misclassification of incident conditions using claims data: impact of varying the period used to exclude pre-existing disease. *BMC Med Res Methodol.* 2013;13:32. doi:10.1186/1471-2288-13-32
- Rosenlund M, Ekström N, Törnblom M, Wintzell V, Stark JH, Titievsky L. Impact of variable look-back periods on the incidence rates of chronic diseases using real world data. *Pharmacoepidemiol Drug Saf.* 2020;29(9):1086-1092. doi:10.1002/pds.5066
- Clemmons NS, Wallace GS, Patel M, Gastanaduy PA. Incidence of Measles in the United States, 2001-2015. *Jama.* 2017; 318(13):1279-1281. doi:10.1001/jama.2017.9984
- Woudenberg T, van Binnendijk RS, Sanders EA, et al. Large measles epidemic in the Netherlands, May 2013 to March 2014: changing epidemiology. *Euro Surveill.* 2017;22(3). doi:10.2807/1560-7917.ES.2017.22.3.30443
- Wettermark B. The intriguing future of pharmacoepidemiology. *Eur J Clin Pharmacol.* 2013;69(Suppl 1): 43-51. doi:10.1007/s00228-013-1496-6
- Caetano PA, Lam JM, Morgan SG. Toward a standard definition and measurement of persistence with drug therapy: Examples from research on statin and antihypertensive utilization. *Clin Ther.* 2006;28(9):1411-1424. doi:10.1016/j.clinthera.2006.09.021
- Mirandola M, Andretta M, Corbari L, Sorio A, Nose M, Barbui C. Prevalence, incidence and persistence of antipsychotic drug prescribing in the Italian general population:

- retrospective database analysis, 1999–2002. *Pharmacoepidemiol Drug Saf.* 2006;15(6):412–420. doi:10.1002/pds.1162
11. Montastruc F, Benard-Laribiere A, Noize P, et al. Antipsychotics use: 2006–2013 trends in prevalence and incidence and characterization of users. *Eur J Clin Pharmacol.* 2018;74(5):619–626. doi:10.1007/s00228-017-2406-0
 12. Brett J, Karanges EA, Daniels B, et al. Psychotropic medication use in Australia, 2007 to 2015: Changes in annual incidence, prevalence and treatment exposure. *Aust N Z J Psychiatry.* 2017;51(10):990–999. doi:10.1177/0004867417721018
 13. Benard-Laribiere A, Noize P, Pambrun E, et al. Trends in incidence use of benzodiazepines and Z-drugs in France from 2006 to 2012: a population-based study. *Pharmacoepidemiol Drug Saf.* 2017;26(2):162–169. doi:10.1002/pds.4123
 14. Forns J, Pottegård A, Reinders T, et al. Antidepressant use in Denmark, Germany, Spain, and Sweden between 2009 and 2014: incidence and comorbidities of antidepressant initiators. *J Affect Disord.* 2019;249:242–252. doi:10.1016/j.jad.2019.02.010
 15. Bramness JG, Sexton JA. The basic pharmacoepidemiology of benzodiazepine use in Norway 2004–9. *Norsk Epidemiologi.* 2011;21(1):35–41. doi:10.5324/nje.v21i1.1423
 16. Su CC, Chia-Cheng Lai E, et al. Incidence, prevalence and prescription patterns of antipsychotic medications use in Asia and US: a cross-nation comparison with common data model. *J Psychiatr Res.* 2020;131:77–84. doi:10.1016/j.jpsychires.2020.08.025
 17. Ravera S, Visser ST, de Gier JJ, de Jong-van den Berg LT. Prevalence, cumulative incidence, monotherapy and combination therapy, and treatment duration of frequently prescribed psychoactive medications in the Netherlands: retrospective database analysis for the years 2000 to 2005. *Clin Ther.* 2010;32(14):2457–2466. doi:10.1016/j.clinthera.2010.12.002
 18. Halfdanarson OO, Pottegård A, Bjørnsson ES, et al. Proton-pump inhibitors among adults: a nationwide drug-utilization study. *Therap Adv Gastroenterol.* 2018;11:1–11. doi:10.1177/1756284818777943
 19. Nam YH, Bilker WB, DeMayo FJ, Neuman MD, Hennessy S. Incidence rates of and risk factors for opioid overdose in new users of prescription opioids among US Medicaid enrollees: a cohort study. *Pharmacoepidemiol Drug Saf.* 2020;29(8):931–938. doi:10.1002/pds.5067
 20. Sørensen AMS, Rasmussen L, Ernst MT, et al. Use of tramadol and other analgesics following media attention and risk minimization actions from regulators: a Danish nationwide drug utilization study. *Eur J Clin Pharmacol.* 2021;77(4):617–624. doi:10.1007/s00228-020-03016-6
 21. Støvring H, Pottegård A, Hallas J. Determining prescription durations based on the parametric waiting time distribution. *Pharmacoepidemiol Drug Saf.* 2016;25(12):1451–1459. doi:10.1002/pds.4114
 22. Kvalitetsdeklaration. Statistik om läkemedel år 2020. March 30, 2021. The National Board of Health and Welfare. <https://www.socialstyrelsen.se/statistik-och-data/statistik/statistikammen/lakemedel/> Accessed December 16, 2021.
 23. Bilaga-Tabeller-Statistik om läkemedel. 2019. The National Board of Health and Welfare. <https://www.socialstyrelsen.se/statistik-och-data/statistik/statistikammen/lakemedel/>. Accessed February 25, 2021.
 24. WTDTTT: Stata module to estimate parameters of the ordinary and reverse Waiting Time Distribution (WTD) by maximum likelihood. Statistical Software Components. [computer program]. 2020.
 25. Altamura AC, Moro AR, Percudani M. Clinical pharmacokinetics of fluoxetine. *Clin Pharmacokinet.* 1994;26(3):201–214. doi:10.2165/00003088-199426030-00004
 26. Støvring H. Selection bias due to immigration in pharmacoepidemiologic studies. *Pharmacoepidemiol Drug Saf.* 2007;16(6):681–686. doi:10.1002/pds.1419
 27. Gardarsdottir H, Heerdink ER, Egberts AC. Potential bias in pharmacoepidemiological studies due to the length of the drug free period: a study on antidepressant drug use in adults in the Netherlands. *Pharmacoepidemiol Drug Saf.* 2006;15(5):338–343. doi:10.1002/pds.1223
 28. Riis AH, Johansen MB, Jacobsen JB, Brookhart MA, Sturmer T, Støvring H. Short look-back periods in pharmacoepidemiologic studies of new users of antibiotics and asthma medications introduce severe misclassification. *Pharmacoepidemiol Drug Saf.* 2015;24(5):478–485. doi:10.1002/pds.3738
 29. Roberts AW, Dusetzina SB, Farley JF. Revisiting the washout period in the incident user study design: why 6–12 months may not be sufficient. *J Comp Eff Res.* 2015;4(1):27–35. doi:10.2217/ce.14.53
 30. Sundell M, Eriksson M, Brynhildsen J, Spetz A-C, Hoffmann M. Misclassification of incidence of menopausal hormone therapy depending on age and length of run-in. Personal communication, in preparation. 2021.
 31. Beaglehole R, Bonita R, Kjellström T. *Basic epidemiology*. WHO; 2004. <https://apps.who.int/iris/handle/10665/43541>
 32. Wallerstedt SM, Wettermark B, Hoffmann M. The First Decade with the Swedish Prescribed Drug Register - A Systematic Review of the Output in the Scientific Literature. *Basic Clin Pharmacol Toxicol.* 2016;119(5):464–469. doi:10.1111/bcpt.12613
 33. The ATC/DDD Methodology. World Health Organization. <https://www.who.int/tools/atc-ddd-toolkit/methodology> Accessed December 17, 2021.
 34. ATC Structure and Principles. WHO Collaborating Centre for Drug Statistics Methodology. https://www.whocc.no/atc/structure_and_principles/ Accessed December 17, 2021.
 35. Bodkergaard K, Selmer RM, Hallas J, et al. Using the waiting time distribution with random index dates to estimate prescription durations in the presence of seasonal stockpiling. *Pharmacoepidemiol Drug Saf.* 2020;29(9):1072–1078. doi:10.1002/pds.5026
 36. Støvring H, Pottegård A, Hallas J. Refining estimates of prescription durations by using observed covariates in pharmacoepidemiological databases: an application of the

reverse waiting time distribution. *Pharmacoepidemiol Drug Saf.* 2017;26(8):900-908. doi:[10.1002/pds.4216](https://doi.org/10.1002/pds.4216)

SUPPORTING INFORMATION

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

How to cite this article: Hoffmann M, Støvring H. Incidence in pharmacoepidemiology—Basic definitions and types of misclassification. *Basic Clin Pharmacol Toxicol.* 2022;130(6):632-643. doi:[10.1111/bcpt.13727](https://doi.org/10.1111/bcpt.13727)