



**HAL**  
open science

# An Overview of Regression Models for Adverse Events Analysis

Elsa Coz, Mathieu Fauvernier, Delphine Maucort-Boulch

► **To cite this version:**

Elsa Coz, Mathieu Fauvernier, Delphine Maucort-Boulch. An Overview of Regression Models for Adverse Events Analysis. Drug Safety, 2023, 10.1007/s40264-023-01380-7 . hal-04382493

**HAL Id: hal-04382493**

**<https://hal.science/hal-04382493v1>**

Submitted on 22 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# An Overview of Regression Models for Adverse Events Analysis

Elsa Coz<sup>1,2,3,4</sup> · Mathieu Fauvernier<sup>1,2,3,4</sup> · Delphine Maucort-Boulch<sup>1,2,3,4</sup>

Accepted: 2 November 2023 / Published online: 25 November 2023  
© The Author(s) 2023

## Abstract

Over the last few years, several review articles described the adverse events analysis as sub-optimal in clinical trials. Indeed, the context surrounding adverse events analyses often imply an overwhelming number of events, a lack of power to find associations, but also a lack of specific training regarding those complex data. In randomized controlled trials or in observational studies, comparing the occurrence of adverse events according to a covariable of interest (e.g., treatment) is a recurrent question in the analysis of drug safety data, and adjusting other important factors is often relevant. This article is an overview of the existing regression models that may be considered to compare adverse events and to discuss model choice regarding the characteristics of the adverse events of interest. Many dimensions may be relevant to compare the adverse events between patients, (e.g., timing, recurrence, and severity). Recent efforts have been made to cover all of them. For chronic treatments, the occurrence of intercurrent events during the patient follow-up usually needs the modeling approach to be adapted (at least with regard to their interpretation). Moreover, analysis based on regression models should not be limited to the estimation of relative effects. Indeed, absolute risks stemming from the model should be presented systematically to help the interpretation, to validate the model, and to encourage comparison of studies.

## Key Points

For the comparison of adverse events between patients, regression models adjusting important covariables may be considered both in observational and randomized controlled trials.

Time-to-event models have been advocated in adverse events analysis. However, models like logistic regression (with rare event corrections) may be considered for rare events.

If possible, the absolute risk from the regression model should be presented systematically because it may help validation and interpretation, particularly in the competing risk settings, and comparison between studies with different follow-up times.

Due to the multiple facets of adverse events data, several risk measures may be relevant to accurately evaluate and compare patients' toxicity profiles.

✉ Mathieu Fauvernier  
mathieu.fauvernier@chu-lyon.fr

<sup>1</sup> Université de Lyon, 69000 Lyon, France

<sup>2</sup> Université Lyon 1, 69100 Villeurbanne, France

<sup>3</sup> Hospices Civils de Lyon, Pôle Santé Publique, Service de Biostatistique et Bioinformatique, 69003 Lyon, France

<sup>4</sup> CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Équipe Biostatistique-Santé, 69100 Villeurbanne, France

# 1 Introduction

Over the past few years, there have been several studies undertaken to depict the situation with regard to the collection, reporting and analysis of drug safety data in randomized controlled trials (RCTs). In a review, the practice regarding the collection, reporting and the analysis of adverse events (AEs) were described as inconsistent and sub-optimal [1]. The CONSORT harm extensions [2, 3] provided guidelines to cover the reporting of harms but have not been sufficiently adopted [1]. With regard to the statistical part of AE analysis, guidelines hardly tackle the question, but focus mainly on collection and reporting [4]. A scoping review of the statistical methods identified several studies designed specifically for AE data, but those methods have rarely been applied [5]. A recent survey conducted by statisticians from academia and industry to understand the current practice [6] identified several barriers that limit the application of those methods. Beyond the barrier of the design and characteristics of the RCT (e.g., the overwhelming number of events and the underpowered sample size for harm outcomes), many participants indicated a lack of guidelines, of awareness of appropriate methods, and of training on the subject.

For treatment comparison, even for randomized designs, it may be interesting to consider multivariate regression models to adjust on important covariables to improve precision and reduce sample size [7–9]. However, they are barely used in practice [10]. In pharmacoepidemiology, i.e., the study of the risks and therapeutic effects of drugs in real-life populations [11], regression modeling is even more relevant, as those studies rely on observational data with no control for confounders, so adjusted analysis may

be necessary to reduce biases. However, due to the multidimensional nature of AEs (e.g., timing, multiplicity, severity) modeling is challenging [12]. For chronic treatments or in oncology, the occurrence of intercurrent events (e.g., death, treatment interruption, treatment discontinuation) is unavoidable even in controlled designs and must be considered during the choice of the model and of the risk measures [13].

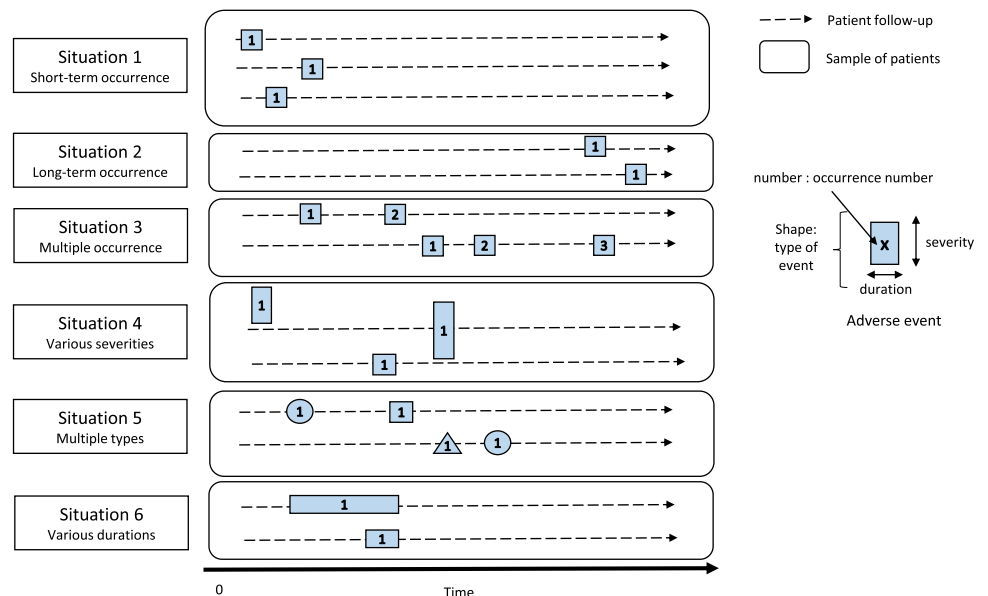
Over the last few years, several authors advocated the use of competing risk methods to avoid bias related to the death in time-to-event outcomes of AE analysis [14, ]. Non-parametric estimation of risk measure was discussed in the context of pharmacoepidemiology, but without considering competing events, and more recently in the context of RCTs [15]. Models for neglected dimensions were proposed like recurrence [16] or severity [17]. To our knowledge, there is no specific overview of which regression model to choose depending on the research question and the data structure. The purpose of this article is to (1) provide an overview of regression models available to compare adverse events between patients, (2) to discuss and give some clues regarding the choice of the model according to the characteristic of the event, the clinical framework, and the objective.

## 2 Models

### 2.1 Towards the Model

The definition of the event(s) and its (their) specificities highly affect the modeling choices. Figure 1 gives a general overview of the characteristics that may considered when identifying a suitable model. The AE might occur early

Fig. 1 Event characteristics driving the modeling choices



or late (situation 1 or 2) after treatment initiation, it might occur more than once (situation 3) or have various severities (situation 4). Furthermore, different AE types may be modelled together to catch a common effect (situation 5) (e.g., different AEs from a same body system [18]). Finally, when available, AE duration may be of interest because long durations may highly affect patients' quality of life (situation 6).

Additional to the event characteristics, the therapeutic pathway of the patients may influence the characteristics of the model. In a chronic disease such as cancer, a patient's follow-up may last several months or years so that during the treatment, many incidents may happen, referred as intercurrent events in the *Addendum on estimands and sensitivity analysis in clinical trials* of the ICH Guidelines [13, 19]. We will mention some of them that usually require the model to be adapted. Figure 2 depicts common successions of incidents as they may be seen during a patient's follow-up.

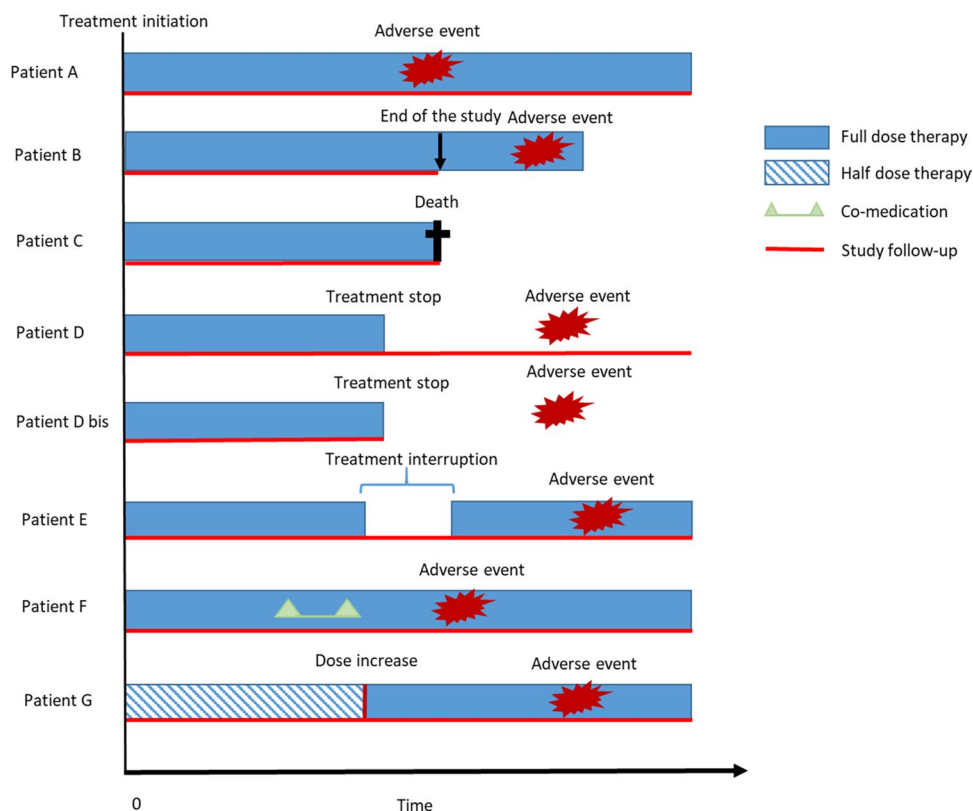
Patient A and B belong to the classical setting of survival analysis, the event of interest (e.g., the AE) occurring at different times and potentially censored (patient B). A first type of intercurrent event is death (Patient C), which obviously prevents the AEs from occurring. When interested in comparing the occurrences at a given time for all patients, the model usually needs to account for a terminating event. Targeting the "direct" effect of the covariable on the occurrence of AEs (i.e., not mediated by the competing event) is much more complicated, as it requires the hypothetical

scenario in which the intercurrent event would not occur. This is not achievable, unlike sometimes making unrealistic/untestable assumptions [20, 21] (e.g., independence between the competing events and the events of interest). We will not address those methods in the article. Patient D discontinues the treatment, which usually reduces or even removes the risk of AEs. Hence, this phenomenon counts as an informative censoring event. Contrary to death, the outcome may not be defined because AE collection may be reduced in frequency or even discontinued as in patient D bis (e.g., if the patient begins another treatment) [22]. Treatment interruption for a given period (Patient E), co-medication (Patient F), or dosage change (Patient G) modify the risk of exposure to AEs and raise important methodological issues, not to mention that collecting information about co-medication from the patients is very difficult and when present may have thousands of modalities.

### 2.2 Single Event Models

The usual practice in AE analysis is to focus on one occurrence of an event (e.g., the first, or the most serious). This section details regression models for a single event outcome. For a formal definition of the outcome and associate risk measure in each case, the reader may refer to Table 1.

Fig. 2 Typical patients' therapeutic pathways



**Table 1** Summary of the models and risk measures of the single-event models section

| Observation  | Model                            | Absolute risk measures  |   | Associated relative risk measure  |
|--|----------------------------------|---|---|---|
|  |                                  | Indicator   | Definition  |   |
| Event occurrence indicator   | $Y$                              | Probability   | $P(Y = 1)$  | Risk ratio <sup>a</sup><br>OR<br>Probability difference   |
| Censoring indicator, time to event                                 | $C, T$                           | Logistic (with/without rare event correction)<br>Time-to-event models (e.g., Cox model) | Hazard $\lambda(t)$<br>Cumulative incidence function $F(t)$   | Use cause-specific hazards:<br>$\lim_{dt \rightarrow 0} \frac{P(t < T \leq t+dt   T \geq t, D \geq t)}{dt}$<br>- Use cause-specific hazard models of each cause of failure to compute the quantity<br>- Fine-Gray model to directly link the covariables to $F(t)$ <sup>b</sup> |
| One measure from an ordered severity scale                         | $G$ where $G = 0$ means no event | Ordinal logistic regression   | Probabilities of severity $G$   | $P(G = g)$<br>Unchanged   |
| Ordered severity scale measures evaluated at several times         | $G_1, G_2, \dots, G_K$           | Mixed model   | Mean grade  | Joint modeling with a time-to-event model on the terminal event may be considered to avoid bias on the covariable effects   |
| Time of occurrence, censoring indicator and ordinal severity scale | $T, C, G$                        | Longitudinal ordinal logistic regression<br>Berridge and Whitehead                      | Probabilities of severity $G$ at time $t$<br>Cumulative incidence function of each severity category<br>Hazard $\lambda(t)$ | $P(G(t) = g)$<br>$P(T < t, G = g)$<br>Cf. time-to-event models  |

$CR$  continuation ratio,  $HR$  hazard ratio,  $OR$  odds ratio

<sup>a</sup>AKA relative risk, risk ratio

<sup>b</sup>The Fine-Gray model is actually based on sub-distributional hazards defined as:  $1 - F(t) = \exp(-\int_0^t h(s)ds)$

## 2.2.1 Logistic Regression

Logistic regression is the most obvious way of relating covariables with a single event. In AE analyses, it may not always be appropriate. Regarding the characteristics we described above, we identified the following situations to be appropriate for this model.

- *Early occurrence* Estimates stemming from the logistic regression (e.g., probabilities or odds ratios [OR]) may be highly biased in presence of censoring (loss to follow-up, administrative censoring) so early AEs may be less sensitive to this issue (situation 1 in Fig. 1) [23]. Moreover, for those events, the times of occurrence may be quite homogenous between patients so that modeling their timing may have limited interest.
- *Rare events* When considering rare events, the logistic regression may be a modeling option. As the event probabilities may be underestimated in those setting [24], it can be combined with useful type of penalization (e.g., Firth's correction) to reduce biases [25]. Variants of Firth's corrections have also been developed to improve the estimates of ORs [26]; they may be particularly useful for analyses of rare but serious AEs [27, 28] with case-control designs [22].
- *Case-control designs* The ORs directly stemming from the logistic regression are known to be interesting for risk evaluation in the case-control study designs, in which the baseline probability is not available.

Odds ratios are often interpreted as probability ratios but this should be done with caution when the probability of an event occurrence is greater than 0.1 [29]. Indeed, in that case, ORs tend to differ from probability ratios and may therefore lead to overestimation of the association between the event and the risk factor. Estimating the risk ratio or risk difference with non-rare outcomes may avoid misleading interpretations, although they require more complex methods (e.g., binomial model) [30].

*Handling terminal intercurrent events* In case of competing events, all those risk measures stemming from the logistic regression provide the total effect of the covariables on the AE occurrence, meaning a combination of both the direct effect on the AE occurrence and the effect mediated by the competing events [31]. That is why indicators based on probabilities are sometimes criticized for not providing information regarding potential differences in follow-up durations between patients [32]. A classic example is the comparison between treatments with the same hazards of AEs in case one of them increases survival or progression-free survival. Considering the probability of event, the conclusion is that the risk of toxicity is larger in the group with increased survival; hence, the explanation is that patients

who die rapidly do not have time to experience AEs in the other treatment arm [19].

## 2.2.2 Time-to-Event Models

To deal with censoring and to describe the occurrence of a single type of AE over time, time-to-event models have been advocated to improve the analysis of safety data [4, 19, 33]. The proportional hazards model (or Cox model) is the most commonly used regression time-to-event model. The following situations (non-exhaustive list) are well suited for time-to-event models:

- *Long-term toxicity analysis with censoring* As long-term cohort analyses are usually affected by censoring (e.g., administrative), so time-to-event models are the only ones to guarantee unbiased estimates of the probability of events in that situation.
- *Most non-recurrent AEs* If the event of interest is not recurrent (e.g., serious events that may lead to treatment discontinuation), building a time-to-event model is usually valuable because in addition to the treatment effect, it provides an interesting description of the risk of toxicity over time (e.g., cumulative hazard or survival in the Cox model). The proportionality assumption may be questionable and systematically checked, particularly when comparing treatments with very different toxicity profiles (e.g., immunotherapy vs chemotherapy). If the proportionality is not valid, flexible models allowing for time-varying effects may be useful [34].
- *Special interest in time-varying covariables such as drug exposure* Another valuable aspect of time-to-event models is their ability to consider important time-dependent covariables, such as exposure. In 2019, Danieli and Abrahamowicz [34] tackled the modeling of drug exposure and treatment interruptions in a time-to-event model to be used in observational studies. They considered a weighted cumulative exposure to deal with both the time elapsed since the last exposure and the cumulative dose. Flexible estimates described the way past exposures affect the hazard of a specific event. This approach is interesting in case of late AEs, potentially triggered by drug accumulation (Fig. 1 situation 2).

Hazard ratios (see Table 1) measure the association between the event and the covariables, although we generally cannot use them to establish causality [35], even with a correctly Cox specified model and a proper randomization at baseline. Indeed, by conditioning on survival, the risk set may be modified over time if the covariable of interest has an actual treatment effect or in presence of an unmeasured covariable [36]. Cumulative probabilities difference or ratio may be considered instead for causal inference. In

both cases, reporting absolute risk measures (e.g., hazards or cumulative hazards, cumulative probabilities) is highly advised [37].

*Handling terminal intercurrent events* As previously mentioned, competing risks are often encountered in AE analysis, particularly in oncology [14] complicating time-to-event analyses and their interpretation [38]. Two main approaches may be used for that type of analysis. The first is to model cause-specific hazards based on the times of occurrence for a single cause of failure (see Table 1). Cox model can be used to perform a regression on the cause-specific hazards. A distinct model for each cause of failure is needed to compute the cause-specific cumulative incidence function. The second approach relies on sub-distributional hazards whose famous associated regression model is the Fine-Gray model [39]. Despite its unclear biological interpretation [40], the strength of the model is to directly link the variables to the cumulative incidence function in a single model [41]. In both cases, covariable effects measures (e.g., difference or ratio) derived from the cumulative probabilities provide the same total effect we previously described for logistic regression [31]. Dealing with the direct effect of the treatment cannot definitely be solved using the cause-specific or the sub-distributional hazards, which have generally no causal interpretation, even in randomized designs [31]. Causation in the competing risks setting is constituting an active fields of statistical research and other estimands may be considered (e.g., survivor average causal effect [SACE]) [21].

### 2.2.3 Models Comparing Severities

The severity of AEs is usually quoted using an ordered categorical or a numeric scale. The well-known National Cancer Institute's Common Terminology Criteria for Adverse Events (CTCAE) uses a grade from 1 (mild) to 5 (death). Then, we may want to compare the impact of a covariable on the severity of an event (situation 4 of Fig. 1). For example, one may be interested in the level of toxicity of a treatment regarding various dose schemes [42] or by the risk factors of severity [43]. A natural option is to conduct an ordinal logistic regression (e.g., proportional odds model or continuation ratio model) [44]. The idea of those models is to associate the covariables with the probabilities of the levels of severities. However, as with the classical logistic regression, the ordinal logistic regression does not account for the timing of the AEs and it may be highly biased in the presence of censoring limiting the situations of application of those models. Hence, other modeling options may be considered according to the aim of the study.

- *The interest is in the evolution of the severity of the adverse events in patients over time* In this case, repeated measurements of the level of severity are collected and

longitudinal models may be considered. A first option is the ordinal logistic regression with a random effect on the patient's level. For example, Augustin et al [45] built a longitudinal proportional odds model on an oral mucositis score, with the cumulative dose and the mouth sites as covariables, to improve the planning of radiation therapy. The second option is to consider the grade (e.g., CTCAE grades) as a repeated measure over time using a linear mixed model [46–48]. The latter approach captures the complete toxicity trajectory of the patients, including the burden of low-grade AEs. It may provide a comprehensive, visual description of the toxicity profile despite the uncertain assumption of grade normal distribution.

*Handling terminal intercurrent events* When the longitudinal outcome and the occurrence of a terminal event (i.e., informative dropout) are correlated, the linear mixed model will lead to biased estimates. To reduce the bias, a joint model may be considered [49], i.e., a mixed model for the longitudinal data and a survival model for the time to death. Shared parameters like random effects link the two outcomes.

- *The interest is in comparing the cumulative probabilities of occurrence for each level of severity over time (potentially the maximum grade per patient)* To handle this issue, Berridge and Whitehead [47] built a two-component model to estimate the occurrence probabilities of AEs according to their severities. One component is a proportional hazard model, used to estimate the all-grade probability of events over time in an unbiased manner, and the second component is an ordinal logistic model which ventilates the probability over the levels of severity.

## 2.3 Recurrent Events Models

Harm studies tend to focus on severe life-threatening AEs (e.g., CTCAE grade 3–4), which are recommended to be systematically collected [4, 50]. As those events may lead to treatment discontinuation, single-event models seem rather suited to them. However, it became increasingly common to include patient reported outcomes (PRO) (e.g., PRO-CTCAE [51]) in clinical trials to complete AE collection because physicians tend to under-report symptoms in terms of frequency and/or severity, compared to patients themselves [52]. Those events may be mild but recurrent (situation 3 in Fig. 1) and may highly affect the quality of life. However, the statistical methods used to deal with these recurrent events have been repeatedly reported as inappropriate (e.g., restricting the analysis to the first or the worst-grade event) in clinical trials [1, 12, 16]. In this section, we discuss the use of some common recurrent events methods



in the context of AEs. The reader may refer to Table 2, for a summary of the models and formal definitions of the outcomes and risk measures.

Previously, in the time-to-event model section, the hazard was defined as the instantaneous probability by unit of time of experiencing an AE in patients who had not experienced an AE previously. Here, the counting process theory is used to extend the notion of hazard to recurrent events. The intensity of AEs is defined as the instantaneous probability by unit of time of experiencing an AE given the history of the process (e.g., the timing of the previous events) [53]. Additionally, ‘rate’ will stand for the instantaneous probability of experiencing an AE. We identified the two following clinical questions that may guide the choice of the recurrent event model.

- *Interest in finding associations between the covariables and the overall occurrence of AEs over time:* The most natural modeling approach is the Poisson and the Negative Binomial model [10]. However, the Andersen-Gill (AG) model [54] has been found to perform better in complex situations and should be preferred (unless the available data are aggregated counts) [55]. The latter is marginal (e.g., based on quantities such as the rate or the cumulative mean) semi-parametric regression on the rate (extension of the Cox model). The AG model may be easily applied with standard statistical software handling the Cox model, by simply rearranging the dataset [53, 56]. A robust estimator of the variance is usually needed to account for correlation between the events of a same subject [57]. As marginal quantities, covariables effect on the rate or on the cumulative mean number of events do have a causal interpretation (no selection of the population over time). Hence the rate is particularly interesting in randomized designs to compare treatment effects [58]. The cumulative mean number of events, that are easier to understand, may be obtained from the rate function.

*Handling terminal intercurrent events* Terminal events may be managed similarly to the single event hazards by considering the modeling of the rate of AEs at a given time, conditioned on the survival from the terminal event at the same time (see Table 2) [53]. Due to this conditioning, the rate-based covariable effect estimates no longer have a causal interpretation [58]. Furthermore, the rate is no longer directly related to the cumulative mean function (it may also depend on the terminating event rate).

- *Interest is to compare the patient’s individual risks of AEs given their number of previous events* The Prentice-Williams-Petersen (PWP) [59] model may be considered in that situation. It is another semi-parametric proportional

model based on intensities and the history of the process being merely summarized by the number of previous events. Hence, it can be seen as a multi-state model. Two formulations of the model are possible depending on the knowledge of the process. The first formulation is defined by the time of the events. The covariable effect on the occurrence of AEs is evaluated over the entire observation period and may be allowed to vary according to the number of previous events. In particular, the event occurrence does not depend on the delay since the previous event. The second formulation relies on inter-times or gap times, i.e., the delay between two subsequent events. In that formulation, the occurrence of an event does not depend on the delay since the beginning of the follow-up (e.g., the beginning of the treatment). Like the AG model, the coefficient estimates may be obtained from a standard statistical software handling the Cox model, by rearranging the dataset and stratifying the number of previous events [53, 60]. As for the hazard, the covariable effects in PWP models do not have a causal interpretation (selection of the population) despite randomization at baseline [61]. Deriving the cumulative mean number of AEs from intensity models can be very complex, as well as can its interpretation [53].

*Handling terminal intercurrent events* The terminal event (e.g., death) may be related to the recurrent process (e.g., serious events). In that case, the joint modeling of the hazard of the terminating event and of the intensity of the process may be necessary (e.g., joint frailty model) [53, 62].

*Comparing severities in recurrent event models* The model of Berridge and Whitehead discussed in the single-event model section was extended to deal with recurrent events by replacing the proportional hazard model by a recurrent model such as PWP [17].

## 2.4 Multi-type Event Models

As discussed in section 2, patients may experience multiple types of AEs (situation 5 in Fig. 1). It is always possible to model them independently but this would result in a loss of power. Modeling AEs jointly seems more attractive. The Wei Lin and Weissfeld (WLW) model [63] is a Cox model extension that is able to handle several types of events (by stratifying the type). It provides a common measure of the variable effect across all types of AEs considered. This comes with an important gain in power but at the cost of the strong hypothesis that covariable’s effects are the same whatever the type of AEs.

*Comparing severities in recurrent event models* An extension of the Berridge and Whitehead model was proposed [58] in which the two components are respectively replaced



**Table 2** Summary of the models and risk measures of the recurrent event models and multi-type events sections

| Observation   | Model  | Absolute risk measures                                     |   | Definition  | In case of a terminating event occurring at time $D$  | Associated relative risk measure |
|---|--|--|---|---|---|----------------------------------|
|   |  | Indicator  |   |   |   |                                  |
| <b>Recurrent events models</b>                              |  |  |   |   |   |                                  |
| Times to events, censoring time                             | $T_1, T_2, \dots, T_k, C$                      | Andersen-Gill + robust estimator of variance               | Rate $\rho(t)$  | $\frac{d\mu(t)}{dt} = \lim_{dt \rightarrow 0} \frac{P(dN(t)=1)}{dt}$ <sup>a</sup> | Rather consider: $\lim_{dt \rightarrow 0} \frac{P(dN(t)=1 D \geq t)}{dt}$                         | Rate ratio                       |
|   |  |  | Cumulative mean function $\mu(t)$   | $\mathbb{E}(N(t))$ <sup>a</sup><br>Directly linked to the rate                    | An additional cause-specific hazard model of the terminating event is required to derive $\mu(t)$ | Mean difference                  |
|   |  | Prentice-Williams-Petersen (PWP)                           | Intensities $\lambda_k(t)^b$  | $\lim_{dt \rightarrow 0} \frac{P(t < T_k \leq t+dt   N(t)=k-1)}{dt}$              | The joint modeling of the hazard and intensity may be needed                                      | Intensity ratio                  |
| Inter-events times, censoring time                          | $U_1, U_2, \dots, U_k, C$                      | PWP gap times formulation (PWP gap)                        | Intensities $\lambda_k(t)^b$  | $\lim_{dt \rightarrow 0} \frac{P(t < U_k \leq t+dt   N(t)=k-1)}{dt}$              |   |                                  |
| Times to events, ordered severity scale measures            | $(T_1, G_1), (T_2, G_2), \dots, (T_k, G_k), C$ | Gebski et al: Ordinal logistic regression + PWP or PWP gap | Cumulative probabilities over time of severity levels<br>Intensity/hazard | $P(T_k < t, G_k = g)$   |   | CR                               |
| <b>Multiple-events models</b>                               |  |  |   |   |   |                                  |
| Times of each individual type of event, censoring indicator | $T_1, T_2, \dots, T_m, C$                      | Wei-Lin-Weissfeld (WLW)                                    | Hazards   | $\lim_{dt \rightarrow 0} \frac{P(t < T_k \leq t+dt   T_k > t)}{dt}$               |   | HR                               |

CR continuation ratio, HR hazard ratio, OR odds ratio

<sup>a</sup>We denote the counting process  $N(t)$ , associated with the recurrent times and the history of the process  $H(t)$

<sup>b</sup>Actually, PWP is a special case of the intensity function, the history of the process being summarized by  $N(t)$ , whose general definition may be written:  $\lim_{dt \rightarrow 0} \frac{P(dN(t)=1|H(t))}{dt}$ .

by a multinomial logistic regression and a recurrent event model (PWP or WLW). One difficulty here is to carefully define mutually exclusive categories of AEs.

### 3 Discussion

In this article, we discussed various regression models for AE outcomes. We considered various dimensions of those events, including severity. Models dealing with severity are more complex (two-component models) and most have been proposed recently. More practical examples to facilitate their interpretation, as well as the implementation of software, would be useful so that they can be used routinely. In most cases, models should be adapted in the presence of a terminal event like death (at least with regard to their interpretation).

Although we extensively searched the literature to illustrate the methods of this overview, this article does not claim to be representative of the actual practice nor to be an exhaustive list of the methods used in that context. We first identified methodological articles dealing with the statistical issues in AE analysis in Pubmed and Google Scholar using keywords: “toxicities”, “adverse events” or “drug safety” with “statistical analysis”. We did not exclude any period of publication. From those articles, we built a list of regression models. To enrich the discussion, we managed to identify articles that apply the models in practice with drug safety data by searching the name of the model with the keywords “toxicities”, “adverse events” or “drug safety” in PubMed and Google Scholar as previously. Hence, this article provides some methodological tools that may suit common situations and answer some clinical questions. Most of the references we provided dealt with the comparison of AEs between treatment arms in RCTs but their usage may be extended to observational studies, like the Qualitop project [64], which motivated this article, and various covariables of interest.

Often, risks measures are used to “map the AE data to a single value” [19] for the purpose of safety evaluation. However, unlike efficacy, AE comparisons may not rely on a single value due to the complex dimensionality of those data. For example, providing both absolute and relative risk measures is commonly advised [37]. For non-parametric estimation, incidence rate is often advised compared to the overall probabilities to account for studies with various follow-up durations (e.g., due to different durations of two treatment arms) [19]. However, by considering the overall cumulative incidence function over time instead of an overall probability, quantities are more comparable. Graphical representations of the absolute risk stemming from the regression model should be done systematically as it may help to validate the adequacy of the model (e.g., comparison with

non-parametric estimates) and to interpret the model, particularly in the presence of competing risks. Moreover, it should facilitate further meta-analyses, mixing studies with various follow-up durations.

All throughout the article, we considered the outcome (AEs) of the models to be clearly defined. However, the number of AEs collected may be huge. For example, the CTCAE has narrowed the keyword field used to describe AEs but its version 4.0 still includes more than 1000 terms. Hence, the analyses and comparisons have to then focus on a small number of events of interest whereas the criteria for their selection are often unclear, ill documented or based on arbitrary rules (e.g., frequencies  $\geq 5\%$ ). Some authors considered grouping AEs according to the body systems [65] but assigning types of AEs into body systems is not always as easy as it may seem and the grouping choices may highly influence the conclusions of the study [18]. Selecting the AEs according to their attributability to the treatment is a more difficult task. In their 2016 recommendations, Lineberry et al did not insist on this kind of selection because of its inherent subjectivity and limited value in clinical trials [4].

One common limitation of all the models we discussed is the reliability of data collection. If mild AEs may be of interest regarding the quality of life of the patients, they are often under-reported by clinicians [52]. Therefore, using Patient Reported Outcomes (PRO) may be more relevant in that situation. Furthermore, we discussed models using severity that may be difficult to collect reliably over the whole follow-up, particularly in observational studies. For severe events, the patient is most likely to come for a consultation or may be hospitalized, so the collection of the AE is close to the time of occurrence. Otherwise, AE collection is usually performed when patient meets clinician for a follow-up visit (e.g., once a month). The time of occurrence is therefore not precisely collected, which leads to interval censoring and modeling issues. Moreover, some AEs may arise and be resolved in between visits (e.g., transient hyperthyroidism during immunotherapy), so they may not be collected (truncation). Hence, comparing the occurrence of such AEs in patients with various visit frequencies may be misleading.

### 4 Conclusion

Comparing the adverse events between groups of patients is a recurrent occurrence with drug safety data. Regression models adjusting on important covariables may be considered in both observational and RCTs. Time-to-event models are advocated for AE analysis; however, the interpretation of those models are complicated because of competition with death or treatment discontinuation. Hence, the absolute risk from the regression model should be presented

systematically because it may help validation and interpretation, particularly in the competing risk settings, and comparison between studies with different follow-up times. Flexible time-to-event models dealing with baseline risks (unlike semi-parametric models) as well as non-linear and time-dependent covariate effects have proven to be useful and should be explored further in this context. Rare events are a recurrent issue in drug safety data and few models may suit rare outcomes. Hence, the logistic regression (with rare event corrections) may be a useful option. Recent articles proposed models accounting for severity; however, their interpretation may be difficult and real-life application should be performed to extend their use.

**Acknowledgments** We thank Jean Iwaz for his careful lecture and corrections. We also thanks the three anonymous reviewers whose comments significantly improve the manuscript.

## Declarations

**Funding** This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 875171. The findings and conclusions in this report are those of the authors and do not necessarily represent the view of the consortium.

**Conflict of Interest** The authors have no relevant financial or non-financial interests to disclose.

**Code Availability** Not applicable.

**Ethics Approval** Not applicable as the study does not involve human participants.

**Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

**Availability of Data And Materials** Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

**Author Contributions** EC completed the literature review, manuscript figures, data summary, and manuscript revisions with input from MF and DMB. All authors were involved in critical revision and approval of the final manuscript drafts.

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

## References

1. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ Open*. 2019;9(2): e024537. <https://doi.org/10.1136/bmjopen-2018-024537>.
2. Ioannidis JPA, Evans SJW, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141(10):781–8. <https://doi.org/10.7326/0003-4819-141-10-200411160-00009>.
3. Junqueira DR, Zorzela L, Golder S, et al. CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials. *BMJ*. 2023;381: e073725. <https://doi.org/10.1136/bmj-2022-073725>.
4. Lineberry N, Berlin JA, Mansi B, et al. Recommendations to improve adverse event reporting in clinical trial publications: a joint pharmaceutical industry/journal editor perspective. *BMJ*. 2016;355: i5078. <https://doi.org/10.1136/bmj.i5078>.
5. Phillips R, Sauzet O, Cornelius V. Statistical methods for the analysis of adverse event data in randomised controlled trials: a scoping review and taxonomy. *BMC Med Res Methodol*. 2020;20(1):288. <https://doi.org/10.1186/s12874-020-01167-9>.
6. Phillips R, Cornelius V. Understanding current practice, identifying barriers and exploring priorities for adverse event analysis in randomised controlled trials: an online, cross-sectional survey of statisticians from academia and industry. *BMJ Open*. 2020;10(6): e036875. <https://doi.org/10.1136/bmjopen-2020-036875>.
7. Tsiatis AA, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Stat Med*. 2008;27(23):4658–77. <https://doi.org/10.1002/sim.3113>.
8. FDA C for DE and. Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products. Published May 25, 2023. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products>. Accessed September 28, 2023.
9. Ishii J, Ohshimo S, Shime N. Potential Confounders for Applying a Novel Sepsis Care Quality Improvement Program. *Crit Care Med*. 2020;48(2):e161–2. <https://doi.org/10.1097/CCM.00000000000004069>.
10. Patson N, Mukaka M, Otworld KN, et al. Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials. *Malar J*. 2020;19(1):119. <https://doi.org/10.1186/s12936-020-03190-z>.
11. Quartey G, Wang J, Kim J. A review of risk measures in pharmacoepidemiology with tips for statisticians in the pharmaceutical industry. *Pharm Stat*. 2011;10(6):548–53. <https://doi.org/10.1002/pst.521>.
12. Cabarrou B, Gomez-Roca C, Viala M, et al. Modernizing adverse events analysis in oncology clinical trials using alternative approaches: rationale and design of the MOTIVATE trial. *Invest New Drugs*. 2020;38(6):1879–87. <https://doi.org/10.1007/s10637-020-00938-x>.
13. FDA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Published 2019. [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf). Accessed 9 Oct 2023.
14. Allignol A, Beyersmann J, Schmoor C. Statistical issues in the analysis of adverse events in time-to-event data. *Pharm Stat*. 2016;15(4):297–305. <https://doi.org/10.1002/pst.1739>.
15. Stegherr R, Schmoor C, Lübbert M, Friede T, Beyersmann J. Estimating and comparing adverse event probabilities in the presence

- of varying follow-up times and competing events. *Pharm Stat.* 2021. <https://doi.org/10.1002/pst.2130>.
16. Hengelbrock J, Gillhaus J, Kloss S, Leverkus F. Safety data from randomized controlled trials: applying models for recurrent events. *Pharm Stat.* 2016;15(4):315–23. <https://doi.org/10.1002/pst.1757>.
  17. GebSKI V, Byth K, Asher R, Marschner I. Recurrent time-to-event models with ordinal outcomes. *Pharm Stat.* 2021;20(1):77–92. <https://doi.org/10.1002/pst.2057>.
  18. Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics.* 2004;60(2):418–26. <https://doi.org/10.1111/j.0006-341X.2004.00186.x>.
  19. Unkel S, Amiri M, Benda N, et al. On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharm Stat.* 2019;18(2):166–83. <https://doi.org/10.1002/pst.1915>.
  20. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci U S A.* 1975;72(1):20–2.
  21. Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA. Separable effects for causal inference in the presence of competing events. *J Am Stat Assoc.* 2022;117(537):175–83. <https://doi.org/10.1080/01621459.2020.1765783>.
  22. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials.* 2012;13(1):138. <https://doi.org/10.1186/1745-6215-13-138>.
  23. Crowe BJ, Xia HA, Berlin JA, et al. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials.* 2009;6(5):430–40. <https://doi.org/10.1177/1740774509344101>.
  24. King G, Zeng L. Logistic regression in rare events data. *Polit Anal.* 2001;9:137–63.
  25. Firth D. Bias Reduction of Maximum Likelihood Estimates. *Biometrika.* 1993;80(1):27–38. <https://doi.org/10.2307/2336755>.
  26. Pühr R, Heinze G, Nold M, Lusa L, Geroldinger A. Firth's logistic regression with rare events: accurate effect estimates and predictions? *Stat Med.* 2017;36(14):2302–17. <https://doi.org/10.1002/sim.7273>.
  27. Southworth H, O'Connell M. Data mining and statistically guided clinical review of adverse event data in clinical trials. *J Biopharm Stat.* 2009;19(5):803–17. <https://doi.org/10.1080/10543400903105232>.
  28. Scalorbi F, Argiroffi G, Baccini M, et al. Application of FLIC model to predict adverse events onset in neuroendocrine tumors treated with PRRT. *Sci Rep.* 2021;11(1):19490. <https://doi.org/10.1038/s41598-021-99048-8>.
  29. Katz KA. The (Relative) risks of using odds ratios. *Arch Dermatol.* 2006;142(6):761–4. <https://doi.org/10.1001/archderm.142.6.761>.
  30. Holmberg M, Andersen L. Estimating risk ratios and risk differences: alternatives to odds ratios. *JAMA.* 2020;324(11):1098–9. <https://doi.org/10.1001/jama.2020.12698>.
  31. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med.* 2020;39(8):1199–236. <https://doi.org/10.1002/sim.8471>.
  32. Siddiqui O. Statistical methods to analyze adverse events data of randomized clinical trials. *J Biopharm Stat.* 2009;19(5):889–99. <https://doi.org/10.1080/10543400903105463>.
  33. O'Neill RT. Statistical analyses of adverse event data from clinical trials: special emphasis on serious events. *Drug Inf J.* 1987;21(1):9–20. <https://doi.org/10.1177/009286158702100104>.
  34. Danieli C, Abrahamowicz M. Competing risks modeling of cumulative effects of time-varying drug exposures. *Stat Methods Med Res.* 2019;28(1):248–62. <https://doi.org/10.1177/0962280217720947>.
  35. Hernán MA. The hazards of hazard ratios. *Epidemiol Camb Mass.* 2010;21(1):13–5. <https://doi.org/10.1097/EDE.0b013e3181c1ea43>.
  36. Martinussen T. Causality and the Cox regression model. *Annu Rev Stat Its Appl.* 2022;9(1):249–59. <https://doi.org/10.1146/annurev-statistics-040320-114441>.
  37. Zavala S, Stout JE. Understanding and communicating risk: assessing both relative and absolute risk is absolutely necessary. *JID Innov Skin Sci Mol Popul Health.* 2022;2(2): 100097. <https://doi.org/10.1016/j.xjidi.2022.100097>.
  38. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med.* 2007;26(11):2389–430. <https://doi.org/10.1002/sim.2712>.
  39. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999;94(446):496–509. <https://doi.org/10.1080/01621459.1999.10474144>.
  40. Andersen PK, Keiding N. Interpretability and importance of functionals in competing risks and multistate models. *Stat Med.* 2012;31(11–12):1074–88. <https://doi.org/10.1002/sim.4385>.
  41. Ozenne B, Sørensen A, Lyngholm ST, Torp-Pedersen C, Alexander GT. riskRegression: predicting the risk of an event using cox regression models. *R J.* 2017;9(2):440. <https://doi.org/10.32614/RJ-2017-062>.
  42. Doussau A, Thiébaud R, Paoletti X. Dose-finding design using mixed-effect proportional odds model for longitudinal graded toxicity data in phase I oncology clinical trials. *Stat Med.* 2013;32(30):5430–47. <https://doi.org/10.1002/sim.5960>.
  43. Kulothungan V, Subbiah M, Ramakrishnan R, Raman R. Identifying associated risk factors for severity of diabetic retinopathy from ordinal logistic regression models. *Biostat Epidemiol.* 2018;2(1):34–46. <https://doi.org/10.1080/24709360.2017.1406040>.
  44. Harrell JFE. Regression modeling strategies. 2nd ed. Springer International Publishing AG; 2015. (2015 édition).
  45. Augustin NH, Kim SW, Uhlig A, Hanser C, Henke M, Schumacher M. A flexible multivariate random effects proportional odds model with application to adverse effects during radiation therapy. *Biom J Biom Z.* 2017;59(6):1339–51. <https://doi.org/10.1002/bimj.201600142>.
  46. Thanarajasingam G, Atherton PJ, Novotny PJ, Loprinzi CL, Sloan JA, Grothey A. Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. *Lancet Oncol.* 2016;17(5):663–70. [https://doi.org/10.1016/S1470-2045\(16\)00038-3](https://doi.org/10.1016/S1470-2045(16)00038-3).
  47. Thanarajasingam G, Leonard JP, Witzig TE, et al. Longitudinal Toxicity over Time (ToxT) analysis to evaluate tolerability: a case study of lenalidomide in the CALGB 50401 (Alliance) trial. *Lancet Haematol.* 2020;7(6):e490–7. [https://doi.org/10.1016/S2352-3026\(20\)30067-3](https://doi.org/10.1016/S2352-3026(20)30067-3).
  48. Wong ML, Gao J, Thanarajasingam G, et al. Expanding beyond maximum grade: chemotherapy toxicity over time by age and performance status in advanced non-small cell lung cancer in CALGB 9730 (Alliance A151729). *Oncologist.* 2021;26(3):e435–44. <https://doi.org/10.1002/onco.13527>.
  49. Rizopoulos D. Joint models for longitudinal and time-to-event data: with applications in R. CRC Press; 2012.
  50. EMA. ICH E19 Guideline - Optimization of Safety Data Collection. Published online 2019.
  51. Basch E, Reeve BB, Mitchell SA, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events

- (PRO-CTCAE). *J Natl Cancer Inst.* 2014;106(9):dju244. <https://doi.org/10.1093/jnci/dju244>.
52. Atkinson TM, Ryan SJ, Bennett AV, et al. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. *Support Care Cancer.* 2016;24(8):3669–76. <https://doi.org/10.1007/s00520-016-3297-9>.
  53. Cook RJ, Lawless JF. The statistical analysis of recurrent events. Springer, New York, 2007. p. 82–9, 218–24, 171–77.
  54. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat.* 1982;10(4):1100–20. <https://doi.org/10.1214/aos/1176345976>.
  55. Jahn-Eimermacher A. Comparison of the Andersen-Gill model with poisson and negative binomial regression on recurrent event data. *Comput Stat Data Anal.* 2008;52(11):4989–97. <https://doi.org/10.1016/j.csda.2008.04.009>.
  56. Ozga AK, Kieser M, Rauch G. A systematic comparison of recurrent event models for application to composite endpoints. *BMC Med Res Methodol.* 2018;18(1):2. <https://doi.org/10.1186/s12874-017-0462-x>.
  57. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J R Stat Soc Ser B Stat Methodol.* 2000;62(4):711–30. <https://doi.org/10.1111/1467-9868.00259>.
  58. Janvin M, Young JG, Ryalen PC, Stensrud MJ. Causal inference with recurrent and competing events. *Lifetime Data Anal.* 2023. <https://doi.org/10.1007/s10985-023-09594-8>.
  59. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika.* 1981;68(2):373–9. <https://doi.org/10.1093/biomet/68.2.373>.
  60. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol.* 2015;44(1):324–33. <https://doi.org/10.1093/ije/dyu222>.
  61. Zhong Y, Cook RJ. The effect of omitted covariates in marginal and partially conditional recurrent event analyses. *Lifetime Data Anal.* 2019;25(2):280–300. <https://doi.org/10.1007/s10985-018-9430-y>.
  62. Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics.* 2007;8(4):708–21. <https://doi.org/10.1093/biostatistics/kx1043>.
  63. Wei LJ, Lin DY, Weissfeld L. Regression Analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc.* 1989;84(408):1065–73.
  64. Vinke PC, Combalia M, de Bock GH, et al. Monitoring multidimensional aspects of quality of life after cancer immunotherapy: protocol for the international multicentre, observational QUAL-ITOP cohort study. *BMJ Open.* 2023;13(4): e069090. <https://doi.org/10.1136/bmjopen-2022-069090>.
  65. Güttner A, Kübler J, Pigeot I. Multivariate time-to-event analysis of multiple adverse events of drugs in integrated analyses. *Stat Med.* 2007;26(7):1518–31. <https://doi.org/10.1002/sim.2637>.