Lag-censoring analysis: lights and shades

Giovanni Tripepi¹, Georg Heinze², Kitty J. Jager³, Vianda S. Stel³, Friedo W. Dekker³,⁴ and Carmine Zoccali¹

¹CNR-IFC/IBIM, Clinical Epidemiology and Physiopathology of Renal Diseases and Hypertension of Reggio Calabria, Reggio Calabria, Italy, ²Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria, ³ERA–EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands and ⁴Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands

Correspondence and offprint requests to: Giovanni Tripepi; E-mail: gtripepi@ifc.cnr.it

ABSTRACT

‘Intention-to-treat’ (ITT) analysis is the recommended approach for the data analysis of randomized clinical trials (RCT). ITT analysis considers patients in the active or in the control arm as originally allocated by randomization, independently of their actual adherence to the assigned treatment. Lag-censoring analysis is a statistical method which takes into account the compliance of patients to the study protocol because the investigator censors a patient when or shortly after he/she stops the treatment being tested. Herein we describe the methodology underlying lag-censoring analysis in general terms and by considering the application of this technique in the analysis of a large RCT in haemodialysis patients, the Evaluation of Cinacalcet Hydrochloride Therapy to Lower Cardiovascular Events (EVOlve) trial. Use and misuse of this technique are discussed.

Keywords: intention-to-treat analysis, lag-censoring analysis

INTRODUCTION

The randomized controlled trial (RCT) is the gold standard study design for testing scientific hypotheses in a clinical scenario. RCTs are conducted to test the efficacy of medical interventions and to collect information about adverse effects of the same interventions [1]. The key feature of standard RCTs is that participants are randomly assigned to undergo the treatment being tested or other alternative therapies. After randomization, the two (or more) study groups are followed up by an identical protocol, the only difference between the care patients receive (clinical tests, outpatient visits, etc.) is intrinsic to the interventions being compared. The fundamental advantage of randomization is that it prevents bias by prognosis and that any differences in known and unknown prognostic factors in the groups being compared are due to chance [1]. Unfortunately, randomized controlled trials often suffer from major problems for measuring efficacy, such as noncompliance, protocol deviations, and patient withdrawals, whether these...
problems are related to side effects or not, and these drawbacks pose important methodological concerns during data analysis and interpretation.

‘Intention-to-treat’ (ITT) is the fundamental approach to the analysis of RCTs. By this approach, all RCTs ‘as randomized’, regardless of the compliance to the treatment they actually receive [1]. In ITT analysis of superiority trials (i.e. RCTs aimed at demonstrating that the experimental treatment is superior to placebo/previous therapy for reducing the risk of a given disease/event), the estimate of treatment effect is usually conservative because in patients who discontinue the experimental drug or start taking a non-study drug or who are submitted to an unplanned co-intervention, all events occurring after treatment discontinuation or after a co-intervention continue to be attributed according to randomization, i.e. to the active or the control arm (Figure 1, upper panel). As a consequence, in ITT analysis, the difference in efficacy is typically smaller than that which we would expect assuming as effective the experimental drug, thus making it harder to reject the null hypothesis of no difference. In contrast, the smaller differences commonly observed in ITT analysis have an obvious opposite effect when testing equivalence or non-inferiority between drugs.

The ITT approach is precious to preserve the balanced distribution of risk factors generated by randomization. However, particularly in RCTs with very long follow-up, estimated treatment effects may be diluted by drug discontinuation, e.g. if patients develop the event of interest during the study long after discontinuation of the active drug. Given the fact that suboptimal compliance to prescribed treatments represents the most frequent violation of study protocol in an RCT, it has been proposed to consider compliance as an inclusion criterion of the ’per protocol analysis’ (PPA). PPA assesses the effect of a given treatment under ideal circumstances taking into account only patients with 100% compliance [1]. However, because compliance is frequently related to study outcomes independently of the treatment under investigation [2], PPA could lead to biased results. For example, problems with data interpretation would occur if noncompliance in the active arm would be related to side effects of the drug and poor prognosis, and would occur more often in the active arm. Consequently, in PPA, the active treatment would seem to be better than it actually would be in day-to-day clinical practice.

Because of the interpretative problems related to ITT and PPA, other techniques have been proposed by epidemiologists and among them lag-censoring analysis can be considered as intermediate between ITT and PPA. Lag-censoring analysis is a method aiming at cleaning the ITT effect of a given drug from the dilution effect resulting from drug discontinuation during follow-up. In this type of analysis, the investigator censors a patient at some time (lag-period) after he/she stops the treatment being tested [3]. As in the classical survival analysis, the proviso for the application of this method is that at any time the patients whose survival times are censored should have the same survival probability as those who are still to be followed in the study. This assumption is clearly violated.

![Figure 1: General representation of a lag-censoring analysis.](image-url)
if the time to drug discontinuation (which is a reason of censoring in the lag-censoring analysis) or the drug discontinuation per se (for example, because of a side effect) is related to the time to event, implying that non-compliant patients (censored) have an incidence rate of the event which could differ from that of patients who remain under observation. If such a situation occurs, censoring at or some time after treatment discontinuation may generate bias [3]. Herein we describe the methodology underlying lag-censoring analysis by using data of a hypothetical trial and then we discuss the application of this technique by using data of the Evaluation of Cinacalcet Hydrochloride Therapy to Lower Cardiovascular Events (EVOLVE) trial [4].

LAG-CENSORING ANALYSIS: A HYPOTHETICAL EXAMPLE

Lag-censoring analysis may be performed at different lag times, i.e. at different times after the drug withdrawal or a scheduled co-intervention which may disturb the interpretation of the trial. An important issue in the lag-censoring analysis is the choice of the lag time which should depend on an accurate knowledge of the pharmacokinetics of the study drugs and their carry-over effects. When the lag time is zero, the patient is censored exactly at the day on which he/she stops the study drug (Figure 1, middle panel) and a lag-censoring analysis with a zero lag time is similar to that of a PPA with the only relevant difference that it does not exclude non-compliant patients from the analysis but considers all data by the same patients collected before drug discontinuation. As a result, an event occurring after the day of drug discontinuation is not considered in the data analysis because it falls in a ‘censored time’. When a lag time of 6 months is selected, the patient is considered under observation for the 6 months following the drug discontinuation and is censored thereafter (Figure 1, bottom panel). If the patient experiences an event within this period, such an event is considered in the data analysis while it is ignored if it occurs after such period. Herein we consider a hypothetical randomized controlled clinical trial testing the effect of a new statin in six patients with hypercholesterolaemia and ischaemic heart disease: three patients were allocated to the active arm (testing the efficacy of the experimental drug A) and the remaining three patients to the control arm (treated with the standard drug B). The study endpoint is all-cause mortality and the study duration is 7 years. In the intention-to-treat analysis (Figure 2, upper panel), two patients died in the active arm over a total of 18.5 person-years of observation time (incidence rate: 10.8 deaths/100 person-years) and three patients died in the control arm over a total of 20 person-years of observation time (incidence rate: 15.0 deaths/100 person-years). The incidence rate ratio (experimental versus control arm) is 0.72 (i.e. 10.8/15.0) indicating that the experimental drug A reduces the incidence rate of death by 28% when compared with the treatment B. However, during the follow-up, three patients discontinued the study drugs: patient 2 in the active group stopped drug A after 5.5 years of treatment whereas patients 4 and 5 in the control

**FIGURE 2:** Graphical representation of intention-to-treat and lag-censoring analyses in a hypothetical randomized controlled clinical trial testing the effect of a new statin in six patients with hypercholesterolaemia and ischaemic heart disease (see also text for more details).
group stopped drug B after 6 years (Figure 2, upper panel). On the basis of previous knowledge based on pharmacokinetics of the study drugs (i.e. the carry-over effect), the investigators of this hypothetical study decided to analyse the data by considering a lag-censoring time of 6 months (i.e. the investigators consider active the effects of study drugs within a 6-month period from the date of discontinuation). By applying this technique, the person-time of patient No. 2 (the only one who discontinued the experimental drug A) reduced from 6.5 years to 6 years and because the death of this patient occurred 12 months after stopping drug (i.e. outside the 6 months lag-censoring interval) (Figure 2, bottom panel) such death will not be considered in the analysis. In contrast, the application of a 6-month lag-censoring time would not affect the total person-time in the control group because patients 4 and 5 died within the 6-month lag-censoring interval (Figure 2, bottom panel). By recalculating the incidence rate of death in the active group, the investigators found 1 patient dying over 18.0 person-years which corresponds to an incidence rate of 5.6 deaths/100 person-years. Thus, the incidence rate ratio (experimental versus control group) in the lag-censoring analysis was 0.37 (i.e. 5.6/15.0) indicating that the experimental drug A reduced by 63% the incidence rate of death when compared with the treatment B. The higher benefit of the experimental drug in the lag-censoring analysis (63% incidence rate reduction, IRR) in comparison to that found in the ITT analysis (28% IRR) is due to the fact that the death of patient 2 was attributed to the active arm even though this patient died as long as 12 months after stopping the experimental drug A (i.e. after a time period in which no carry-over can be attributed to treatment A). In this case, lag-censoring analysis resulted in a stronger effect of the drug than ITT analysis (RR = 0.37 instead of 0.72). In the example we reported the lag time is in between the time of drug discontinuation and the time of death. In general, the lag time should be specified in the study protocol and selected based on external knowledge (of pharmacokinetics, as assumed here). Therefore, it should be unrelated to the observed times between drug discontinuation and death.

USE AND MISUSE OF LAG-CENSORING ANALYSIS

When applying lag-censoring analysis, the investigator should be aware that this technique presents some potential drawbacks. First of all, (i) we cannot be completely sure that an event after drug discontinuation is completely unrelated to the previous use of the same medication (long-term harmful effects can occur); (ii) the assessment of the compliance and the exact date in which the drug was discontinued could be problematic; (iii) if a co-intervention is applied more frequently to patients in the placebo arm (as it occurred with statins in the 4D study, Ref. [5]), a higher number of patients are censored in the same arm and this phenomenon unbalances the total person-time of the exposure to the study interventions (active drug/placebo) making it artifactualy shorter in the placebo arm; (iv) the time to drug discontinuation or the drug discontinuation per se should be unrelated to the time to event to avoid bias; (v) if lag-censoring analysis was not in the original protocol, readers would feel you fooled them with this ‘trick’ only to arrive at a significant result. Thus, any conclusion about the results of an RCT should be drawn by ITT approach (an analysis which answers to the question about effectiveness in real situation) and the per protocol and lag-censoring analyses (which answer to the question about efficacy in ideal situation) could serve only as secondary analyses. In general, all limitations and assumptions related to the per protocol analysis also applies to the lag-censoring analysis.

LAG-CENSORING ANALYSIS IN CLINICAL TRIALS IN NEPHROLOGY: THE EVOLVE STUDY

Clinical trials in dialysis patients are a daunting undertaking. These patients are at exceedingly high risk for all-cause and cardiovascular death and nephrologists participating in clinical trials very often perceive it as an ethical obligation to start treatments with drugs of proven efficacy for other clinical conditions. Due to the unique pathophysiological and clinical risk profile of end-stage kidney disease (ESKD) these patients are prejudicially excluded from clinical trials [6]. Furthermore, ensuring compliance in the dialysis population is a notorious problem [7, 8]. As for co-interventions in clinical trials in dialysis patients, 23% of patients in the group assigned to atorvastatin discontinued treatment in the 4D study and as many as 15% of patients in the placebo in the same trial started a non-study statin [5]. The same problem was registered in the AURORA study (29% of patients discontinued the drugs in both the study arms) [9]. In the recent EVOLVE trial, 23% of patients in the placebo arm started non-study Cinacalcet, a figure double than that observed in the active arm (11%). Parathyroidectomy, a co-intervention impacting upon parathormone, i.e. the actual target of Cinacalcet, was applied to 2 and 8%, respectively, of patients in the active and control arm of the study. Furthermore, due to side effects, the investigators had to stop Cinacalcet in a substantial proportion of patients in the active arm of the trial. All these co-interventions may seriously disturb the appreciation of the true effect (i.e. the effect under ideal circumstances) of Cinacalcet on clinical outcomes. Due to these problems, the EVOLVE study was perhaps the most difficult study to analyse among clinical trials performed so far in this population. It is because of these difficulties that the team of biostatisticians that analysed the EVOLVE study applied for the first time the method to control for interventions (i.e. lag-censoring analysis) that might have disturbed the appreciation of the effects of Cinacalcet (stopping Cinacalcet in the active arm, start of non-study Cinacalcet in the control arm or parathyroidectomy in both study arms) on the outcome of this study (a composite of death, myocardial infarction, hospitalization for unstable angina, heart failure, or a peripheral vascular disease).

The results of the primary—intention-to-treat—analysis of EVOLVE were substantially negative [hazard ratio (Cinacalcet versus placebo): 0.93, 95% CI: 0.85–1.02, P = 0.11]. In the secondary analysis, using lag censoring, the survival times were
censored at different lag times (0, 3, 6, 9, 12, 18 months) to clean the intention-to-treat effect estimate from the dilution effect deriving from Cinacalcet withdrawal, the use of non-study Cinacalcet and parathyroidectomy. In EVOLVE, known reasons for discontinuing a study drug before an end point included parathyroidectomy, initiation of commercially available Cinacalcet, and kidney transplantation. Censoring of data at 6 months after study-drug discontinuation suggested a beneficial effect of Cinacalcet on the occurrence of the primary end point [hazard ratio (Cinacalcet versus placebo): 0.85, 95% CI: 0.76–0.95, P = 0.003] with a relative risk reduction of 15%, a figure more than double than that observed in the intention-to-treat analysis (7%). The EVOLVE investigators specifically adopted a 6-month lag time as the anticipated duration for any effect of altered mineral metabolism on extraskeletal calcification [4]. However, one could also argue that in EVOLVE the remarkable improvement in benefit revealed by lag-censoring analysis is difficult to interpret. Starting non-study Cinacalcet and parathyroidectomy were much more common in the placebo arm (30%) than in the active arm (14%). In the intention-to-treat analysis, starting non-study Cinacalcet in placebo treated patients is bound to dilute any underlying benefit of this drug. However, such a dilution effect might have been compensated or even over-compensated by the fact that parathyroidectomy was a more frequent censoring event in placebo treated patients, particularly so in the young (<65 years) subgroup (see EVOLVE Supplementary Appendix, Figure S10) [4], i.e. the age subgroup with a relatively lower hazard rate of the combined end point. Indeed, more parathyroidectomies may translate into a misleading reduction of the total person-time in the placebo arm because such a reduction mainly derives from the subgroup which was less likely to experience the combined end point. Since the person-time is the denominator of the incidence rate, censoring patients because of parathyroidectomy is likely to have induced an under-representation of patients at the lowest risk in the placebo group thus leading to a possible overestimation of the protective effect of Cinacalcet. Of course, such potential bias is not only related to the assessment of the hazard ratio (Cinacalcet versus placebo) but also to the Kaplan–Meier survival curves estimation [10]. In line with this reasoning, the results of the lag-censoring analysis in EVOLVE should be considered with great caution and only the results of the primary analysis should be considered for regulatory purposes.

The lag-censoring analysis was also used by the EVOLVE investigators in a secondary analysis of the same trial for assessing the potential beneficial effect of Cinacalcet for reducing the risk of fractures in haemodialysis patients [11]. Clinical fractures (i.e. fractures ascertained on clinical grounds) were observed in 255 out of 1935 patients (13.2%) who were randomized to the control arm and in 238 out of 1948 patients (12.2%) who were randomized to the active arm. In the ITT analysis, the hazard ratio of fractures was 11% lower in Cinacalcet treated than in untreated patients (HR: 0.89, 95% CI: 0.75–1.07) but this difference did not attain the statistical significance (P = 0.22). Using a pre-specified lag-censoring analysis (6 months lag time), a significant effect of Cinacalcet was found and patients consuming this drug had a hazard ratio of fractures which was 28% lower than that of untreated patients (HR: 0.72, 95% CI: 0.58–0.90, P = 0.009). A lag-censoring analysis carried out by considering a zero lag time (i.e. patients were censored exactly at the time of parathyroidectomy, transplantation, or starting commercial Cinacalcet) provided almost identical results (HR: 0.71, 95% CI: 0.58–0.87, P = 0.001). However, for the same reasons listed above, lag-censoring analysis is likely to have amplified the beneficial effect of Cinacalcet on the risk of fractures by artificially shortening the total person-time particularly in young patients of the placebo arm, i.e. in the subgroup of patients which was less likely to experience a fracture.

**HOW TO ACCOUNT FOR DRUGS DISCONTINUATION IN RCTs: THE INVERSE PROBABILITY OF CENSORING WEIGHTED ANALYSIS**

The inverse probability of censoring weighted (IPCW) analysis is one of the most used statistical techniques for dealing with drug discontinuation in RCTs [12]. This type of analysis was used in the EVOLVE trial to account for the potential bias which may have been introduced due to informative dropout (discontinued study drug prior to having an end point event or ending study). By this method, the EVOLVE investigators assigned a weight for each patient based on the inverse of the predicted probability of continuing to receive a study drug conditional to a series of predictors of compliance. Therefore, heavier weights were given to patients who did not drop out but had similar characteristics to those who did. By using the IPCW technique, the use of Cinacalcet associated to 33% reduction in the risk of the primary end point (HR: 0.77, 95% CI: 0.66–0.88, P < 0.01). The IPCW technique provides valid results only if the exchangeability assumption is ascertained [12]. This assumption only holds when patients who are censored because of drug discontinuation have the same prognosis of patients who remain under observation, an issue which was not reported in the EVOLVE trial. Other more sophisticated statistical techniques for dealing with drug discontinuation are described elsewhere for interested readers [13].

Treatment withdrawal and trial contamination (i.e. the administration of Cinacalcet to patients on placebo) are not unique to EVOLVE in RCTs in ESKD patients. Along with regulatory issues related to the registration and the use of drugs before that the same drugs are demonstrated to be efficacious for the reduction of meaningful clinical end points, specific, predefined strategies for limiting the disruptive effects of drug discontinuation and trial contamination on the interpretation of the study results should be specifically developed for future trials in ESKD patients.

In conclusion, lag-censoring analysis is a statistical method which could serve as a secondary analysis of RCTs with a relevant number of patients who discontinue the study drugs. It should be noted that in a RCT it makes a difference whether an active drug is discontinued or if patients in the placebo arm start using the active study medication non-experimentally. Lag-censoring analysis does not distinguish between these two cases. In some situations though, it may be sensible to apply lag censoring only in the placebo arm, and use the unmodified ITT
principle for the active arm. This analysis strategy still attributes any negative side effects or consequences of poor compliance of patients on experimental active drug to the active treatment, while dilution of the treatment effect by contamination of the placebo arm with non-experimental active drug use is avoided. Furthermore, the validity of the lag-censoring analysis depends on the rigorous assessment of several assumptions including the notion that the time to drug discontinuation or the drug discontinuation *per se* should be unrelated to the time to event to avoid bias. Unfortunately, reasons for drug discontinuation are multifactorial and some of the reasons tend to be harder to ascertain than others. The plurality of the causes tends to make it difficult for investigators to decide whether the application of lag-censoring analysis is justified or not.

**ACKNOWLEDGEMENTS**

This study is part of the Syskid project which is supported through European Union’s FP7, Grant agreement number HEALTH-F2-2009-241544.

**CONFLICT OF INTEREST STATEMENT**

C.Z. received honoraria for lectures by Abbvie, Amgen, Genzyme, Roche and Shire and G.T. by Abbvie. Nothing to declare for G.H., V.S.S. and K.J.J. F.W.D. received unrestricted research grants from Amgen, Baxter and Genzyme. This paper has not been published previously in whole or part, except in abstract format.

**REFERENCES**


Received for publication: 18.11.2014; Accepted in revised form: 26.2.2015