quinten health

Master of Public Health

Master de Santé Publique

Analyzing Recurrent Events in Multiple Sclerosis: A Review of Statistical Models with Application to the MSOAC database

David HERMAN

Class and year of the Master:
MPH - Advanced Epidemiology and Biostatistics - Europubhealth 2022-2024

Location of the practicum:
Quinten Health – Paris, France

Professional advisor:
Quentin Pilard - Quinten Health

Academic advisor:
Pascal Crepey - EHESP

Acknowledgements

I would like to express my sincere gratitude to MSc. Quentin Pilard for his support and interest since the beginning of my internship. His knowledge, motivation, and attention to detail have had a great impact on the development of my thesis and have strengthened my enthusiasm for continuing in the field of biostatistics. I also extend my thanks to Dr. Julien Tanniou for his valuable suggestions and constructive feedback have significantly contributed to my thesis. During my time at Quinten Health, I have had the privilege of working alongside highly skilled professionals who have made me feel welcomed and part of the team.

I am also grateful to the two institutes that have academically trained me (EASP and EHESP) throughout my Europubhealth+. This program has opened doors for me in the field of public health and will undoubtedly be one of the most enriching experiences of my life, with professors, health professionals and friendships that I will always remember.

Finally, these last two years would not have been possible without the unconditional support of my family and my group of friends. A special mention to my parents, Juan Carlos and Mirtha, and my brother Damián, who have always been my main pillar and have supported me in every decision I have made, prioritizing my education, personal goals, and happiness above all else.

Contents

List of acronyms

AG: Andersen Gill

AIC: Akaike Information Criterion

ARR: Annualized Relapse Rate

BIC: Bayesian Information Criterion

CDISC: Clinical Data Interchange Standard Consortium

CDP: Confirmed Disability Progression

CDP12: 12-Week Confirmed Disability Progression

CI: Confidence Interval

CP: Counting Process

CoxPH: Cox Proportional Hazard

DALYs: Disability-Adjusted Life Years

EDSS: Expanded Disability Status Scale

EMA: European Medicines Agency

GEE: Generalized Estimating Equations

GT: Gap Time

HR: Hazard Ratio

IQR: interquartile range

LL: log-Likelihood

LWA: Lee Wei Amato

LWYY: Lin Wei Yang Ying

MCF: Cumulative Mean Function

MS: Multiple Sclerosis

MSM: multi-state models

MSOAC: Multiple Sclerosis Outcome Assessments Consortium

NB: Negative Binomial

PWP: Prentice Williams Peterson

PPMS: Primary Progressive Multiple Sclerosis

QPoisson: Quasi-Poisson

RCT: Randomized Clinical Trial

RR: Rate Ratio

RRMS: Relapsing-Remitting Multiple Sclerosis

SD: Standard Deviation

SE: Standard Error

TT: Total Time

WLW: Wei Lin Weissfeld

List of Tables

List of Figures

Abstract

Introduction: Patients with multiple sclerosis (MS) are susceptible to experience recurrent events of disability progression and relapses. Many studies still focus on analyzing MS events with traditional methods such as Cox Proportional Hazard, Poisson, and logistic regression that either ignore subsequent events or fail to account overdispersion and dependency between events. Therefore, considering recurrent event methods may improve treatment development.

Objective: To conduct a literature review to identify recurrent event methods in the context of MS and apply them on the MS Outcome Assessments Consortium (MSOAC) to provide recommendations for MS research.

Methods: A literature review was conducted to identify main methods, which were then summarized based on their classification and main characteristics. These methods were applied to the MSOAC database to evaluate the effect of the disease course on the number of confirmed disability progression (CDP) and the Annualized Relapse Rate (ARR).

Results: A total of 54 articles were included in the literature review, identifying 9 main recurrent event models. The most documented were the Andersen-Gill, Prentice Williams and Peterson and Frailty models. Marginal models may be recommended in experimental studies over conditional approaches, while event-specific models are accurate for estimating overall or event-specific effects in patients with one or more events. Random effect models are suited for studies with patient heterogeneity. In the MSOAC database, recurrent events have provided more precise estimates than traditional methods. Common and event-specific estimates for CDP and ARR were consistent across models.

Conclusion: This study provides methodological guidance for health researchers to select and implement appropriate methods in recurrent event analyses. The model choice may vary depending on the research study and different factors. Researchers should prioritize recurrent event methods in their statistical plans to avoid information loss and improve the precision of estimated effects.

Key words: Recurrent events, Count data, Proportional Hazards Models, Multiple Sclerosis

Resume

Titre : Analyse des événements récurrents dans la sclérose en plaques : Revue de la littérature et application à la base de données MSOAC

Introduction : Les patients atteints de sclérose en plaques (SEP) sont susceptibles d'expérimenter des événements récurrents de progression du handicap et de rechutes. Toutefois, de nombreuses études appliquent des méthodes statistiques usuelles telles que la régression de Cox, de Poisson ou logistique. Or, ces méthodes ignorent soit les événements ultérieurs ou bien ne tiennent pas compte de là sur dispersion et de la dépendance entre les événements. L'utilisation de méthodes dédiées aux évènements récurrents serait plus pertinente et pourrait donc améliorer significativement le développement de futur traitement.

Objectif : Mettre en place une revue de la littérature afin d'identifier les méthodes d'analyse des événements récurrents dans le contexte de la SEP et les appliquer au Multiple Sclerosis Outcome Assessments Consortium (MSOAC) dans le but de fournir des recommandations claires pour les cliniciens du domaine.

Méthodes : Une revue de la littérature a été menée afin d'identifier les principales méthodes d'analyse des événements récurrents. Ces méthodes ont été résumées et classifiées selon leurs principales caractéristiques. Ensuite, elles ont été appliquées sur la base de données du MSOAC afin d'évaluer la récurrence de progression confirmées de l'incapacité (CDP) et la récurrence de rechute (ARR) en fonction de la typologie de la SEP (récurrente-rémittente ou forme secondairement progressive).

Résultats : Au total, 54 articles ont été inclus dans la revue de la littérature, identifiant 9 modèles d'événements récurrents. Les modèles les plus documentés sont ceux d'Andersen-Gill, de Prentice Williams et Peterson et à fragilité partagée. Les modèles marginaux peuvent être recommandés dans les études expérimentales par rapport aux approches conditionnelles, tandis que les modèles spécifiques à l'événement sont précis pour estimer les effets globaux ou spécifiques à l'événement chez les patients ayant un ou plusieurs événements. Les modèles à effets aléatoires conviennent aux études avec hétérogénéité des patients. Dans la base de données MSOAC, les événements récurrents ont fourni des estimations plus précises que les méthodes traditionnelles. Les estimations communes et spécifiques à l'événement pour le CDP et l'ARR étaient cohérentes d'un modèle à l'autre.

Conclusion : Cette étude fournit des recommandations aux cliniciens spécialisés dans la sclérose en plaque pour les aider à sélectionner et à mettre en œuvre les méthodes dédiées aux

évènements récurrents. Le choix du modèle peut varier en fonction de l'étude de recherche et de différents facteurs. Les cliniciens devraient favoriser ce type de méthode en tant que critère de jugement principal lors de l’élaboration de futures études cliniques afin de maximiser l’information et donc d’améliorer la précision des effets estimés.

1. INTRODUCTION

1.1. Multiple Sclerosis and associated clinical outcomes

Multiple sclerosis (MS) is a leading cause of non-traumatic neurological disability in young adults. It affects the brain and spinal cord, resulting in symptoms like blurred vision, weak limbs, tingling sensations, dizziness, and fatigue (1). With over 2 million global prevalent cases and an annual incidence rate of 2.1 per 100,000 person-year, in 2016, MS led to 18,932 deaths and 1,151,478 Disability-Adjusted Life Years (DALYs) (2).

The disease is classified into three types (or disease courses): the relapsing-remitting (RRMS), secondary progressive (SPMS) and primary progressive (PPMS). The RRMS and SPMS are the most frequent and are characterized for the presentation of repeated events of isolated attacks (relapses) or progressive disability (3–6). The RRMS, which is the most prevalent (85% of total cases), is defined by relapses of new or increasing neurological symptoms. The SPMS type involves an initial relapsing-remitting phase followed by a progressive worsening of symptoms, independent of relapses (7,8). Although MS prevalence has increased partly due to improved survival, there is still no cure, and the disease-modifying therapies with earlier diagnosis are the main interventions to reduce symptoms and slow the worsening of the disease (1,9,10).

Relapses and disability progression are the main outcomes to evaluate disease progression in MS (10). The latter is often measured by the changes in the Expanded Disability Status Scale (EDSS), which ranges from 0 to 10, with higher score indicating more severe disease (11,12). These changes are frequent primary endpoints in clinical trials and can be derived by measuring confirmed disability progressions (CDP) during follow-up (8,11,13–15). Relapses, defined as new or worsening clinical signs or symptoms lasting at least 24 h without fever, are mainly measured by the Annualized Relapse Rate (ARR) in RRMS clinical trials (6,10,16). The combination of these endpoints describes more accurately a heterogeneous disease with a variety of subtypes (4,10).

1.2.Traditional statistical methods in MS research.

Most of researchers often analyze repeated events of disability progression and relapses in MS using well-established methods (6,17,18). Their choice varies according to the nature of the outcome variable. For time-to-event data, survival models are used, with the Cox proportional hazards model (CoxPH) being a common choice, focusing on the time to the first event (8). For count data, Poisson regression is often implemented to examine the number of events over time. For the continuous and categorical results, linear regression and logistic regression models are used, respectively (6).

Nevertheless, when these traditional methods are applied to recurrent event data, various issues may arise. The CoxPH can be inefficient because the information following the first disability progression or relapse is ignored (6,17–20). For Poisson, the problem is that recurrent event data (i.e. MS relapses) often present over-dispersion with variance larger than the mean, which results in parameters with lack of precision and statistical significance can be overestimated (6). Moreover, in Poisson regression, the within-subject correlation is not correctly accounted (19). Both linear and logistic regression may be inappropriate for analyzing this type of data, due to specific assumptions and outcome redefinition, which may lead to waste of information and impact the statistical power and parameter estimation (6,17).

1.3.Recurrent events methods

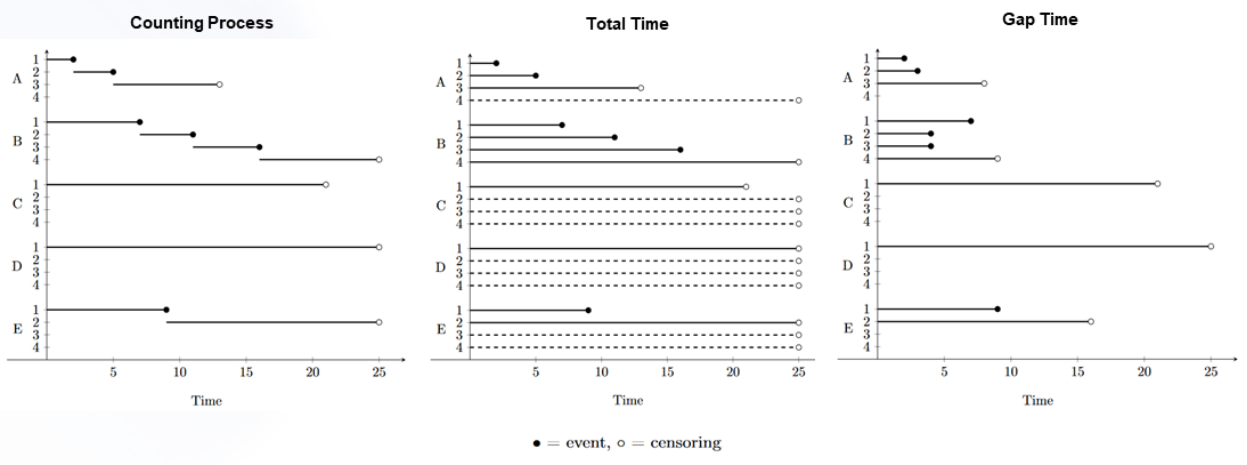
1.3.1. Definition

In view of these limitations, a variety of counts and survival statistical methods, called recurrent events methods, have been developed for the correct estimation of repeated events while accounting for its lack of independence (4,6,8,19,21–23). Recurrent events refer to the repeated occurrence of the same type of event over time for the same individual, such as hospitalizations, asthma attacks and multiple sclerosis relapses or disability progressions (8,19,22). Although recurrent events can also consider terminal events (i.e. death), in MS the likelihood of terminal events is low and the interest of recurrent events methods rely on non-terminal events (8,24,25).

Recurrent events are characterized with five main components: 1) time scale defined as calendar time (the time measured from the time origin) or gap time (time is reset to zero after each event); 2) risk interval which defines when an individual is at risk of having an event along a given timescale; 3) risk set or the number of individuals who are at risk at a given point in time; 4) an event-specific or common baseline hazard; and 5) the way of handling the within-subject correlation, which can be conditional, marginal or random effects (8,22,26).

The risk interval is a key concept in recurrent events that allows to define whether a model is either marginal or conditional, and can be categorized as counting process, total time, and gap time (Figure 1) (8). The counting process uses calendar time as time scale but considers also left truncation (i.e. delayed entry). For example, individual B is at-risk for the first event between [0, 7) and for the second, third and fourth event during [7, 11), [11, 16) and [16, 25), respectively. An individual is not at-risk for a specific event before previous event has been observed. Total time corresponds to the time from time since the beginning of the observation in the study. For instance, individual B is at-risk for the first event in the time interval [0, 7), for the second event during [0, 11), for the third event during [0, 16) and for the fourth event during [0, 25). On the other hand,

gap time is the time from the prior event. The individual B is assumed to be at risk for the first event during [0, 7) and for the second, third and fourth event during [0, 4), [0, 4) and [0, 9).



Source: modified from Bühler (8).

Figure 1. Risk intervals illustration recurrent event data for four hypothetical individuals.

Another important concept in recurrent event methods is the type of intervention effect measure obtained (27). Methods can be based on cumulative events (number of events by end of study), event rate[1] (number of events per unit time), time to event (time to successive events) and gap time (times between successive events are accounted). Event rate models yield a rate ratio (RR) as estimation measure (i.e. Poisson and negative binomial models), and models based on time-to-event (i.e. CoxPH) provide a hazard ratio (HR) as measure of effect.

1.3.2. Classification

The classification of recurrent event methods may vary depending on the approach considered by the author (18,20,22,28,29). However, this is mainly based on risk interval and within-subject correlation definitions as well as the type of intervention effect measure (8,26,27).

Based on risk intervals, recurrent events methods are classified as conditionals or marginals models (8,26,28). Conditional models use a full specification of the recurrent event process through the event history (8). The conditioning on the event history can be, either through a random effect term (i.e. frailty term) or through time-dependent covariates (i.e. symptoms and event counters) (8,29). In marginals, full specification of the recurrent event process may remain unspecified and models are focused on marginal parameters as the dependence structure is not

[1] Some models also known as count-based models such as Poisson or Negative Binomial models. These are statistical models used to analyze count data.

of interest (8,19). These can be classified as marginal hazard models or mean/marginal rate models depending on the intervention effect measure obtained, which are the hazard ratio and the event rate, respectively (8,27).

1.3.3. Applicability of recurrent events

Recurrent event methods are designed to address several research questions (Table 1). Although they can be adapted to individual analyses, characterization and association studies, recurrent event are predominantly focused on treatment effect estimation in clinical trials (27). Common research questions are related to the measuring of the intervention effect or in estimating the number of prevented cases in the experimental group.

The use of recurrent event methods provides significant advantages in both statistical analysis and public health. Its application can lead to more accurate hypothesis testing, as these methods appropriately accounts for the occurrence of repeated events over time (19). If correlations between events are ignored, the null hypothesis may be incorrectly rejected (18). In addition, recurrent events methods allow to estimate direct and indirect effect of an intervention and to gain considerable statistical power, leading to smaller sample sizes, shorter follow-up, or both (24,30).

Within the MS context, the use of recurrent events methods is widely suggested by the scientific community, and it was recently recommended by the European Medicines Agency (EMA) (31). Their use may enhance greater precision and better understanding of the disease burden. In particular, they may allow for the consideration of about 10% and 23% of disability events in RRMS and PPMS trials, respectively, that are often excluded (21,24,32–34). In addition, enable to supply the real target for treatment both from a patient, provider (i.e. economic) and societal perspective, examine the potential benefit of a new treatment, to better characterize the prognosis of patients, and to facilitate access to potential new therapies (6,8,21–24,32–34).

Table 1. Research questions and objectives in recurrent events methods with examples based on MS.

Domain	Question	Objective	Example
Individual states	What are the factors associated with relapses/death in patients?	Understanding and describing individual event processes	Evaluation of factors on the risk of illness-relapse/death process in a patient
Characterization	What factors contribute to variations in the frequency and severity of recurrent MS relapses among individuals?	Identifying and characterizing variation across a population of processes	Risk factors or protective factors identification for severe or frequent relapses
Group comparations	How many events does the treatment prevent, on average? Does the treatment decrease the event number/rate over the study period? What is the intervention effect on the number of higher-order events? What is the effect of intervention on the number of subsequent events?	Comparing groups of processes	Treatment effect evaluation in a clinical trial and estimation of disability progression reduction rate
Association	How factors like age, treatment type, and lifestyle influence the likelihood of getting a new relapse?	Determining the relationship of fixed covariates, treatments, and time-varying factors	Influential risk factors of MS patients with relapses

Source: author's own creation with objectives and questions modified from (22,27)

1.4. Research problem

Despite all the mentioned advantages and several reviews suggestions, there remains a lack of consensus agreement on which methods and models should be applied in MS to answer specific research questions (4,6,8,20,35). Most researchers still rely on traditional statistical techniques for analyzing their data with recurrent events, possibly due to lack of awareness, consensus, no external validation of recent models or limited knowledge for its application because of their general complexity (6,10,17,28,29,33). The differences between methods and its advantages and disadvantages remain unclear for most health professionals (33).

Since it may affect the development of optimal treatment strategies and resource allocation in health systems, there is a need to bridge the gap between theoretical advancements in statistical methodology and their practical application in MS research. Thus, this study aims to conduct a well-structured literature review to identify available models, and subsequently, to summarize and apply these models to generate recommendations for improving future research in MS.

1.5.Objectives

General:

To conduct a comprehensive review with an application of recurrent event methods in a MS database to provide recommendations for general use in MS research.

Specific objectives:

1. Identify recurrent event methods in MS and other low-rate mortality diseases through literature review.
2. Describe the advantages and weaknesses of the identified approaches.
3. Apply recurrent event methods to the Multiple Sclerosis Outcome Assessments Consortium (MSOAC) placebo database.

This master's thesis was conducted during a six-month internship at Quinten, a French consulting company in Paris, France. The master's thesis was developed in at the company while the student was also trained in code quality with different tools (Python, VS Code, Docker, GitLab and Amazon Web Service) as well as participating in different meetings. This work arises from the detected gap in the company to make available the statistical models available in recurrent events and its application in future projects. It also seeks to make the information available to the entire research field for the development of research methodologies in chronic diseases.

2. METHODS

2.1. Literature review

A literature review was conducted on PubMed in accordance with Cochrane recommendations for searching and selecting studies (36). Prior reference articles related to the topic were considered and manual searches were also carried out to search other relevant literature sources (i.e. conferences and thesis).

Research query code included the following key words for title and abstract search [tiab] and associated mesh terms [Mesh]: “recurrent event”, “count data”, “chronic disease” and “statistical

models". The search was not limited to MS to be more sensitive and to obtain a broad range of models applied to chronic diseases that could be also used in MS. The selection criteria targeted documents published during the last 20 years (from 2004 onwards), in English and with access to the abstract and full text. The exclusion criteria were based on the type of statistical model mentioned. Records were classified in "recurrent event methods", "time to first event methods", "other methods" (i.e. logistic regression) and "not mentioned a method or statistical model in the abstract". Articles that were not related to the type of study of interest were also excluded.

Extraction was performed in February 2024. After the first step of filtering records, the screening of articles was carried out following the mentioned exclusion criteria. In addition, full-text records were added from Research rabbit & Connected papers[2] by using all the screened articles as main source of searching. Manual searched, and reference articles based on recurrent events were also included. In the final eligibility phase, non-relevant articles were removed. Selected articles were characterized based on type of research item (i.e. review, original research article) and type of recurrent event model addressed.

2.2. Information summarization

Relevant statistical models identified from articles were summarized and organized following Buhler classification, with minor adaptations (8). In this summary, only models for non-terminal events were considered as they are predominant in chronic diseases. For instance, multi-state models (MSM) were not finally considered as these models are mainly used for composite endpoints (illness-death models) in which MS and many other low mortality rate diseases do not apply (37). Each model was outlined with the following components: definition, primary assumptions, advantages and disadvantages, and its relevance in MS.

2.3.Application

2.3.1. Data source and study population

Data from the Multiple Sclerosis Outcome Assessments Consortium (MSOAC) Placebo Database was used in this study (38). It was established in December 2012 by the National Multiple Sclerosis Society and the Critical Path Institute (C-path).

The MSOAC comprises 2465 placebo arms patients from 9 MS clinical trials, conducted primarily from 2000 to 2013 with participants from different countries. All data were acquired prospectively except for some variables (e.g., time since diagnosis), which are typically determined

[2] Both are AI-based tools for exploring and finding research article suggestions based on similarity.

retrospectively. The database does not contain standard-of-care or active comparator data. Data followed the Clinical Data Interchange Standard Consortium (CDISC) standard, widely used in clinical trials and is available to researchers who submit and are approved for its use (38).

For this study, 1313 participants with complete data of the end of the follow-up were included to evaluate the effect of disease course (RRMS diagnoses or SPMS diagnoses) on the of CDP and relapse recurrences. The follow-up was from the first day of the study until the last day of follow-up documented for each patient. Patients were right censored at the end of the study and censoring was assumed to be non-informative[3].

2.3.2. Main outcome measures

The primary endpoint was the confirmed disability progression (CDP) through the EDSS score, defined as an increase from baseline of at least 1.0 EDSS point (baseline EDSS ≤5.5) or ≥0.5 points (baseline EDSS >5·5 points). This definition also considered EDSS progression events regardless of the subsequent EDSS scores (Figure S6) (11). The secondary endpoint was the ARR, calculated as the total number of confirmed relapses that occurred between baseline and the end of the follow-up in years (9,39).

The main exposure variable was disease progression, defined as patients with a diagnosis of RRMS or SPMS type. Results were adjusted by age (years), sex (male, female), race (white, no white) and time since diagnosis (years) as these are known factors associated with MS disease (5). Socio-economic factors were not available in the MSOAC database and country of origin was not considered due to the high proportion of missing values (>50%).

2.3.3. Statistical methods

Descriptive statistics were used to summarize baseline characteristics. Continuous variables were presented as mean and standard deviation (SD) or median and interquartile range (IQR), depending on the distribution. Categorical variables were presented as total number and percentage.

Regarding statistical modelling, several survival and event rate models were applied based on the literature review results. A CoxPH and a Poisson model were used as baseline models. As recurrent event models, extended versions of the CoxPH model (i.e. AG, PWP and WLW) and other event rate models (i.e. NB and Quasi-Poisson) were applied. Hazard Ratio (HR) and Event

[3] Censored patients have the same risk for relapses as those who are not censored (6).

Rate Ratio (RR) were the effect measured estimated. All presented models were adjusted on established confounding factors (40).

Across similar models, the goodness of fit was assessed and compared using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log-likelihood (20,41). These results are shown in supplementary material (Table S11). Proportional hazard assumption for the baseline CoxPH model was tested. No imputation was done for missing data to the selected variables, except for start of observation, which was defined as time t = 1 to all participants.

Data structure for fitting models were constructed based on models' characteristics (8,18,42). Analyses were performed in R (Version 2023.12.0) and syntax for fitting each model with appropriate packages was provided in the supplementary material (Table S11) (8,19,42,43).

2.3.4. Ethical aspects

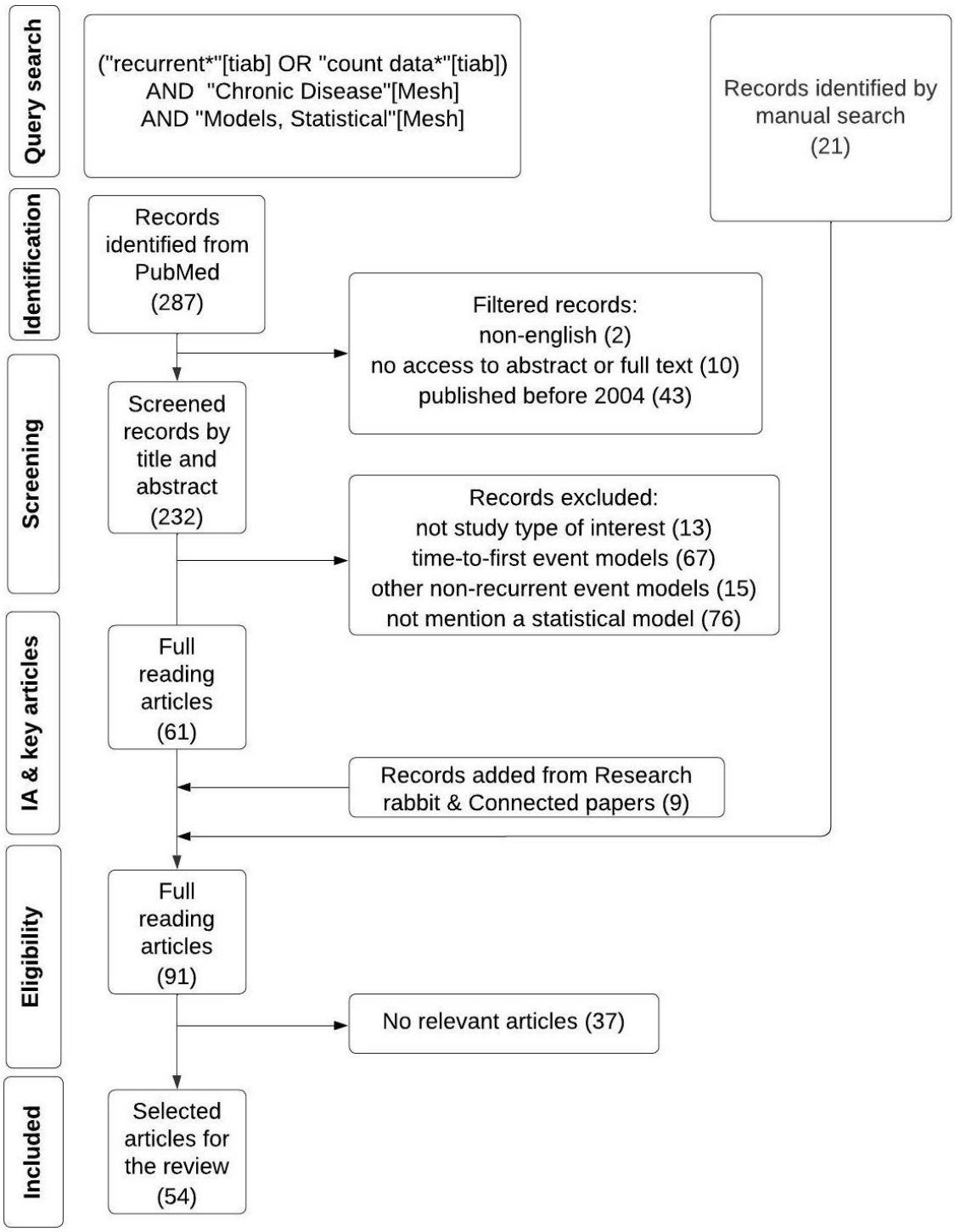
Main author received approval from C-path to use the MSOAC data. All data are fully anonymized and de-identified and ethical considerations have been met.

3. RESULTS

3.1. Literature review findings

A total of 287 articles were found in PubMed. Other 21 publications were added from the manual search and 9 using AI-recommended search tools (Figure 2). After filtering and removing records that not fulfilling the inclusion criteria, 54 publications were selected. These included original statistical model application reviews (n = 28), research articles (n = 14), methodological articles (n = 8), a book (n = 1), a guideline (n = 1), a thesis (n = 1), and a conference article (n = 1).

A total of 9 main statistical models for non-terminal events were identified in the literature review (Figure S5). The most applied model was the Andersen-Gill (AG) model in 29 publications. Other well-documented models were the Frailty models (n = 19), Prentice-Williams-Peterson (PWP-TT or PWP-GT) model (n = 18), Negative Binomial (NB) model (n = 15), and Wei-Lin-Weissfeld (WLW) model (n = 15). Less documented models in recurrent event were the Poisson or its variance-corrected versions (n = 7), the marginals Lin-Wei-Yang-Ying (LWYY) (n = 6), Lee-Wei-Amato (LWA) model (n = 4) and the Partially conditional rate-based (PCRB) model (n = 2).



Source: Author creation with Lucid ®

Figure 2. Attrition flowchart for literature review strategy.

3.2. Classification and description of main statistical models

Following a modified classification from Buhler (8), statistical models for recurrent events are described and detailed information of each model are summarized in the Table 2. This modified version retains aspects proposed by different reviewed authors, allowing the most relevant statistical models to be presented in a simplified approach (8,26,27). It is important to mention that several modifications can be found in a specific model (i.e. Frailty), in which arguments can be adjusted to generate a modified version of the general formula (44).

A. Conditional models

i. Conditional based models

Poisson regression

The general Poisson regression is often used to model recurrent event data in the form of counts (6). If survival time follows an exponential distribution, Poisson is equivalent to a parametric survival model (4).

This model has three main assumptions that generally affects its application in recurrent events: a) its variance should be equals to the mean (equidispersion assumption); b) the successive events occur independently at a constant rate among all patients in each subgroup; and c) the event count should follow a Poisson distribution (6,45).

As recurrent event data often exhibit over-dispersion, the estimated parameters may lack precision and the statistical significance will be overestimated (6). This over-dispersion is often corrected by the inflation of the variance with in the Quasi-Poisson or GEE Poisson models (4,6). The latter models are frequently used in recurrent events as correct the standard errors of the effect estimates and the GEE Poisson also accounts for heterogeneity (46).

Negative Binomial (NB)

The NB regression provides an improved model for recurrent events data compared with Poisson regression (46). This model assumes that each individual has their own underlying event rate over time, but may differ across individuals when it includes a random component reflecting the uncertainty about the true rates (6,8,47). The number of events for each individual follows a Poisson distribution but the expected number is allowed to vary across patients according to a gamma distribution (24,46,48).

Because the NB model with random effect accounts for subject heterogeneity and provides a valid mean rate estimates with a relative easy application, it is frequently suggested in analyses of recurrent events (8,24,46)

Andersen-Gill (AG)

The AG model is a generalization of the CoxPH model and is considered a reference model for recurrent events (21,24,48). It is based on the idea that a patient who experiences a non-fatal event remains in the set of patients at risk (common baseline hazard), recording several events for an individual patient, by considering the interval times and estimating a global estimate effect (8,18,24).

An important assumption of this model, which is often not fulfilled in many diseases, is that the events themselves do not affect the patient's risk of other events, which means that the events are assumed to be independent (30). This common baseline assumption may limit its use in practice, as it can significantly underestimate the overall effect (4,6,24,45,49). For relaxing this assumption, is possible to introduce in the model measured covariates that induce correlation among events for each individual (19). If not possible, a robust sandwich covariance can be used to anticipate correlations among the observations (19,20).

The AG model is usually indicated for analyzing data when all dependence between subsequent events is mediated through time-varying covariates (6,19,22).

Prentice, Williams and Peterson (PWP)

The general PWP model is a stratified AG model, with a separate fitted model for each event (26,45,50). This model accounts for the dependence between repeated events by stratifying on the number of preceding events (event-specific baseline hazard), which means that a subject is assumed not to be at risk for a subsequent event until a current event has terminated (8,19,49).

There are two variations of PWP that depend on the type of risk interval: the PWP-TT (total time)[4], which is similar to the AG model but evaluates the effect of a covariate by stratifying the event and, the PWP-GT (gap time), that is similar PWP-TT, but assumes all events start at the time since the previous event (19,49).

[4] PWP-TT can be also found as *PWP-CT* as it actually uses counting process formulation (26).

An inherent problem with PWP is that, in practice, it may be necessary to limit the data to a specific number of recurrent events, as the risk set may be small for the strata and event-specific estimates may be unreliable (19).

PWP-TT models can be appropriate when risk of event increases with subsequent event while the PWP-GP can be of interest when the distribution of event per subject is small or prediction of time to the next event is of interest (18).

ii. Random effect models

Frailty model

The frailty model is an extension of the CoxPH model, in which the hazard function depends on an unmeasured random variable or "frailty term" that induces a correlation (18,24). Frailty models assume that the within-subject correlation is due to the tendency of some individuals being more prone to develop recurrent events as compared with others because of some unobserved or unmeasured factors (8,19,20). Estimate interpretation is similar to the mixed models in longitudinal data (18). The most used is a shared frailty model with random effects assumed to follow a gamma distribution with mean equal to one and unknown variance (19).

A common limitation comes from the assumption that the effect estimates are constant over the time which can be invalid in experimental designs. Moreover, the potential presence of time-dependent covariates may affect the analyses since it is complex to separate time-dependent effects from unobserved heterogeneity (51). When random effects are large, a small number of events may be adequate for implementation, otherwise a larger number is needed (9,19).

Frailty models are indicated when there exist heterogeneous susceptibility to the risk of recurrent events that cannot be explained by covariates alone (18,19).

B. Marginal models

i. Marginal hazard models

Wei-Lin-Weissfeld (WLW)

This marginal model is based on a total time scale, uses event-specific baseline hazards and a semi-restricted risk set (8). The main assumption of this model is the individual is simultaneously at risk for all events (6).

The WLW approach present some limitations. When it is used in ordered events, this model tend to overestimate the effect (6,26,52). In addition, as other stratified models, data should be limited to an adequate number of recurrent events to still get precise estimates (8).

Compared to other models in which events happen successively, the WLW method is well suited to multi-type event data in which the natural order of the repeated events is no predictable (8). This method preserves the randomization groups in clinical trial with unordered events, providing clear interpretation of the between-group differences (52,53).

Lee-Wei-Amato (LWA)

This is a less known marginal model that only differs to the WLW in its common baseline hazard and unrestricted risk set (8). LWA model was developed originally with the aim to study clustered data (42).

This model often provide bias estimates effect because it allows a subject to be at risk for several events simultaneously (26). However, due to its unrestricted risk set, the LWA is suitable for clustered data such as siblings, where it can be assumed that the baseline hazard is the same, and the beginning of risk of the event is the same within the cluster (26).

ii. Marginal means/rate models

Lin-Wei-Yang-Ying (LWYY)

LWYY can be seen as analogue to the AG model, but with less stronger assumptions as it allows arbitrary dependence structure between recurrent events and a varying rate function over time (8,19). The target estimate measure is the rate ratio and the inference is based on a robust sandwich variance estimator (21).

No assumptions regarding the baseline rate function or the dependence structure between the recurrent events are required, but it typically assumes recurrent event process and time to censoring to be independent (given covariates) (21).

This is one of the most used rate-based models for recurrent event analyses in absence of terminal events or with negligible mortality in clinical trial designs (8).

Partially conditional rate based (PCRB)

This is a less used marginal rate-based model. The PCRB model with a common effect is similar to the LWYY model, but additionally adjusts for event-specific baseline rate functions (21). The event-specific PCRB gives effect estimates of the among all subjects for each stratum as the conditional event-specific PWP (21).

Marginal methods based on PCRB are increasingly used for the analysis of recurrent results in recent years in experimental designs (8).

Table 2. Summary table of main models used in recurrent event analyses.

Models	Risk interval	Risk set	Baseline hazard	Within-individual correlation	Strengths	Weaknesses	References
Poisson	TT	na	na	Non-accounted	• Simplicity in results interpretation.	• Events are independent. • No timing of events. • Ignores heterogeneity. • Estimated parameters may lack precision and significancy may be overestimated	(4,6,8,19,24, 43,45,46)
Quasi- or GEE Poisson	TT	na	na	Random effects (GEE)	• Simplicity in results interpretation. • Variance correction. • Accounts heterogeneity (GEE)	• No timing of events. • Ignores heterogeneity (Quasi-Poisson)	(4,6,8,46,54)
NB*	TT	na	na	Random effects	• Accounts for heterogeneity. • Provides valid mean rates and reliable estimate of RR. • Statistical flexibility.	• No timing of events. • Assumes a constant rate function that can be not true in practice.	(6,8,24,46–48,55)
AG	CT	Unrestricted	Common	Conditional	• Make full use of the data. • Useful when all dependence between subsequent events is mediated through time-varying covariates. • Broader application for the estimation of the overall effect. • Robust variance account for correlation within events.	• Events are assumed to be independent. • The common baseline does not apply to many diseases. • May underestimate the overall effect. • May lose the benefits of randomization in experimental designs. • Proportionality hazard assumption may be not true in practice. • Ignores heterogeneity.	(4,6,8,19,21, 24,30,45,49, 56,57)

PWP-GT	GT	Restricted	Event-specific	Conditional	• Useful if a renewal happens after each event. • Recommended when the distribution of event per subject is small. • Prediction of time to the next event is of interest. • Robust variance account for correlation within events	• Does not coincide with the nature of most chronic diseases. • The risk set may become small for the strata. • Event-specific estimates may be unreliable. • Ignores heterogeneity.	(6,8,18,19,45,49)
PWP-TT	CT/TT	Restricted	Event-specific	Conditional	• Useful if there is more interest in estimate the separate risk for each event. • Allows evidence about risk factors associated with different strata. • Robust variance account for correlation within events	• The risk set may become small for the strata. • Event-specific estimates may be unreliable. • Ignores heterogeneity	(4,6,8,18,19,45,49)
Frailty	TT	Semi-restricted	Common	Random effect	• Accounts for within subject correlation and heterogeneity • Useful when unmeasured heterogeneity cannot be explained by covariates alone (individuals with different risk).	• Event rates are constant over the follow-up time. • It becomes complicated to separate time-dependent effects from unobserved heterogeneity. • Sample size needs to be considered for stable estimates.	(18–20,24,50,51)
WLW	TT	Semi-restricted	Event-specific	Marginal	• Useful when the order of the events is not predictable, and the subject is simultaneously at risk for all events (compatibility with unordered events).	• Limited to a few events in health. • Subjects are at risk for all events even in those with only one event. • May overestimate the treatment effect. • Difficult to interpret global estimate effects.	(4,6,8,26,50,52,53,58)

LWA	TT	Unrestricted	Common	Marginal	• Useful for clustered data where it can be assumed that the baseline hazard is the same.	• Limited to a few events in health. • Subjects are at risk for several events simultaneously. • May provide biased estimates effects.	(4,8,55)
LWYY	TT	Unrestricted	Common	Marginal	• Useful when there are no time dependent covariates, and the dependence structure is not of interest. • Allows for causal interpretation. • Rate functions are more interpretable than the hazard. • Less stronger assumptions. • It is valid regardless of the shape of the baseline rate function.	• Does not specify dependence structures among recurrent event times within a subject. • Assumption of independent time to censoring and recurrent event is often violated when the observation is stopped after a predefined number of events.	(8,13,19,21,58,58,59)
PCRB	TT	Restricted	Event-specific	Marginal	• Adjusts for event-specific baseline rate functions. • Rate functions are more interpretable than the hazard.	• The risk set may become small for the strata. • Event-specific estimates may be unreliable.	(8,21)

Abbreviations: Generalized Estimating Equations (GEE), Negative Binomial (NB), Andersen Gill (AG), Prentice-Williams-Peterson (PWP-GT), Prentice-Williams-Peterson (PWP-TT), Wei-Lin-Weissfeld (WLW), Lee-Wei-Amato (LWA), Lin-Wei-Yang-Ying (LWYY), Partially conditional rate-based (PCRB), na (not apply), TT (total time), CT (Counting time), GP (Gap time); * NB model with random effect.

3.3. Application to the MSOAC database

3.3.1. Implementation of recurrent events in R

3.3.1.1. Data set layout

ARR endpoints from the MSOAC were used to exemplify the data structure appropriate for fitting all recurrent event models without terminal events addressed in this study.

In classical count-based models such as Poisson regression, each participant contributes one record, which includes the number of events as main outcome and total length of follow-up (Table 3). For Data layout 1, *Id* (modified for illustrative purposes) represents the unique identifier of the individual, *Disease Course* the type of disease, *Count* the total number of observations during the total *Length Time* since study start.

Table 3. Data layout 1: count-based models for three individuals.

Id	Disease Course	Count	Length Time
1	SPMS	0	667
2	RRMS	1	813
3	RRMS	2	534

Identification number (id); Number of total events (Count); Relapse Remitting MS (RRMS) and Secondary Progressive MS (SPMS).

On the other hand, data organization in survival-based models is essentially differentiated on the time scale used (CT, TT and GT) and the occurrence of successive events (42,60). In the conditional and marginals rate base models, a subject is assumed not to be at risk for a subsequent event until the current event has finished. However, in the marginal hazards models (WLW and LWA) this assumption is different, as each participant is simultaneously at risk for the occurrence of any event from the beginning of the study (8).

For illustrate these differences, in the Table 4 is displayed the data layout for conditionals and marginals rate-based models. The variable *Id* is a unique patient identifier. *Tstart* and *Tstop* represent the time interval of each observation while *Tgap* the difference of time between observations. Event (0 or 1) represents whether an event occurs at the end of the time interval. If an event has been observed at time *Tstop*, *Event* is equal to 1. If Tstop is a right censoring time *Event* is equal to 0. *Sevent* records the event sequence for each patient, which is necessary for stratified models such as PWP and WLW. *Nevents* summarizes the total number of events experienced by a patient during follow-up. *Disease course* defines the patient's group that in this study is RRMS or SPMS. In data layout 2, patients without relapses have only 1 line, whereas

patients with at least one event have more than 1 line, with the last line corresponding to the time of right-censoring.

Table 4. Data layout 2: conditionals and marginals rate models for three individuals.

Id	Tstart	Tstop	Tgap	Event	Sevent	Nevents	Disease Course
1	1	667	666	0	1	0	SPMS
2	1	268	267	1	1	1	RRMS
2	268	813	545	0	2	1	RRMS
3	1	131	130	1	1	2	RRMS
3	131	401	270	1	2	2	RRMS
3	401	534	133	0	3	2	RRMS

Identification number (id); time start (tstart); time stop (Tstop); time gap (tgap); event sequence (Sevent); Number of total events (Nevents) Relapse Remitting MS (RRMS) and Secondary Progressive MS (SPMS).

Data layout 3 for WLW and LWA should be arranged as each participant have the same number of entries (Table 5). That means that each *id* has many lines as the maximum number of events that could be observed. In this example, maximum number of *Sevent* was defined *Sevent* <= 3 for illustrative purposes.

Table 5. Data layout 3: WLW and LWA marginals hazards models for three participants.

Id	Tstop	Event	Sevent	Disease Course
1	667	0	1	SPMS
1	667	0	2	SPMS
1	667	0	3	SPMS
2	268	1	1	RRMS
2	813	0	2	RRMS
2	813	0	3	RRMS
3	131	1	1	RRMS
3	401	1	2	RRMS
3	534	0	3	RRMS

Identification number (id); time stop (Tstop); event sequence (Sevent); Relapse Remitting MS(RRMS) and Secondary Progressive MS (SPMS).

3.3.1.2. Packages and R commands

Survival, Stats, frailtyEM and *Reda* packages from R were used for conducting the analyses. The *survival* package provides the functions for conducting CoxPH extended models and testing proportional hazard assumptions while the *Stats* package offers the generalized linear models function for count-based models. The *frailtyEM* was only employed to get the frailty variance*. Reda*

was used for graphs and estimation of the mean cumulative function. R commands for fitting models are shown in the supplementary material (Table S11).

3.3.2. Database overview

3.3.2.1. Baseline characteristics and outcome descriptions

From the total participants, 60.2% (721/1313) had a RRMS diagnosis while the remaining 39.8% (522/1313) had a SPMS diagnosis. The mean age of the RRMS population was significantly lower than the SPMS population. Most of participants were female and white in both groups. The SPMS group exhibited a significant longer mean duration since diagnosis, averaging 15.5 years. Additionally, their mean EDSS score was higher (mean = 5.4). Overall, the median total follow-up for all participants was 540 days (Q1: 337, Q3: 673).

Table 6. Baseline characteristics of study population by disease course type.

Variable	RRMS, N = 791	SPMS, N = 522
Age (years)		
Mean (SD)	36.5 (9.1)	49.4 (8.1)
Sex		
Male	240 (30%)	193 (37%)
Female	551 (70%)	329 (63%)
Race		
White	687 (88%)	503 (96%)
No white	98 (12%)	19 (4%)
Missing	6	0
Time since diagnosis (years)		
Median (Q1-Q3)	2.0 (1.0, 5.0)	14.5 (7.8, 22.0)
Missing	0	26
EDSS overall		
Median (Q1-Q3)	2.00 (1.50, 3.50)	6.00 (4.50, 6.50)

Relapse Remitting MS (RRMS) and Secondary Progressive MS (SPMS).

Regarding outcome description for both groups, there were a total of 497 CDP and 406 relapses at the end of the follow-up. The 25.1% (330/1313) and 18.8% (248/1313) of participants had at least 1 CDP or relapse respectively. Around 12% of participants had more than one CDP and relapses, with a maximum of 6 CDP and 9 relapses observed (Table 3). The proportion of events per group was considerably similar for both outcomes.

Table 7. Number of confirmed disability progression (CDP) and relapses by disease course.

Variable	RRMS, N = 791	SPMS, N = 522
Number of CDP		
0	526 (66%)	299 (57%)
1	168 (21%)	162 (31%)
2	69 (8.7%)	53 (10%)
3	15 (1.9%)	7 (1.3%)
4	10 (1.3%)	1 (0.2%)
5	2 (0.3%)	0 (0%)
6	1 (0.1%)	0 (0%)
Number of relapses		
0	500 (63%)	407 (78%)
1	171 (22%)	77 (15%)
2	73 (9.2%)	27 (5.2%)
3	29 (3.7%)	8 (1.5%)
4	9 (1.1%)	3 (0.6%)
5	5 (0.6%)	0 (0%)
6	3 (0.4%)	0 (0%)
9	1 (0.1%)	0 (0%)

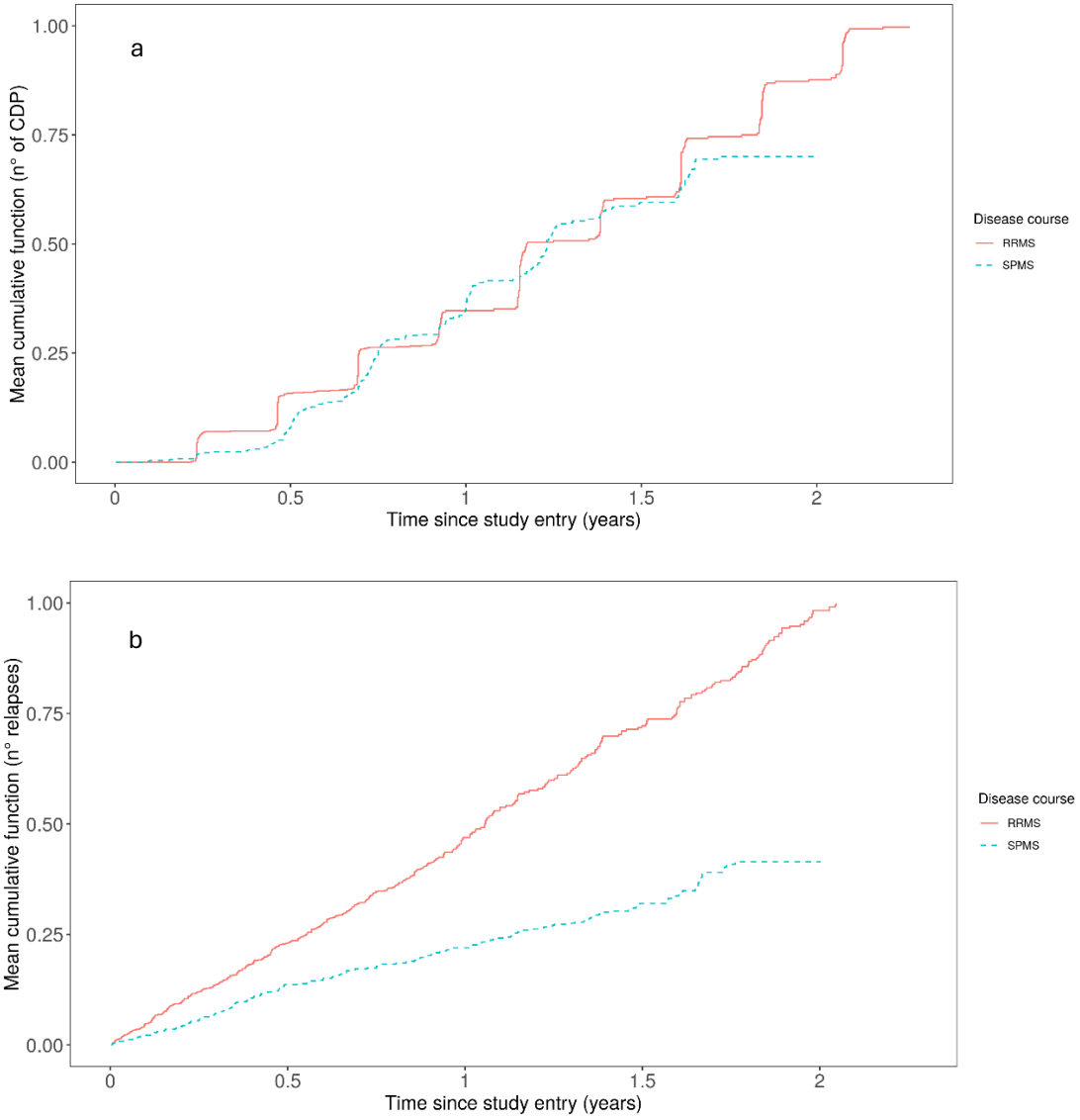
Relapse Remitting MS (RRMS) and Secondary Progressive MS (SPMS).

3.3.2.2. Cumulative number of CDP and relapses.

The mean cumulative function (MCF) is recommended for illustrating recurrent events, as the Kaplan–Meier curves only look at the first event. The MCF represents the average number of cumulative events experienced by an individual in the study at each point in time since the start of follow-up (43).

The figure 3 shows how the cumulative rates of CDP and relapses in RRMS and SPMS groups varies over time. The x-axis and the y-axis represent the time since study entry in years and the average number of events that an individual had experienced during follow-up, respectively. Participants with a RRMS or SPMS diagnosis had in overall, a similar average of CDP during time of follow-up. However, after 1.5 years can be observed a higher average number of CDP in those with RRMS diagnosis. For relapses endpoints, the cumulative relapse rate was lower in the SPMS group. The MCF value was close to 0.9 at year 2, which means that RRMS patients experienced,

on average, 0.9 relapses over the first 2 years of follow-up in the study compared with 0.4 relapses for SPMS participants. It can be also seen that both cumulative mean of CDP and relapses are approximately linear in both groups, suggesting constant relapse rates over time.



Relapse Remitting MS (RRMS) and Secondary Progressive MS (SPMS).

Figure 3. Mean cumulative function of (a) confirmed disability progression and (b) relapse events by disease course.

3.3.3. Effect of disease course on CDP and ARR.

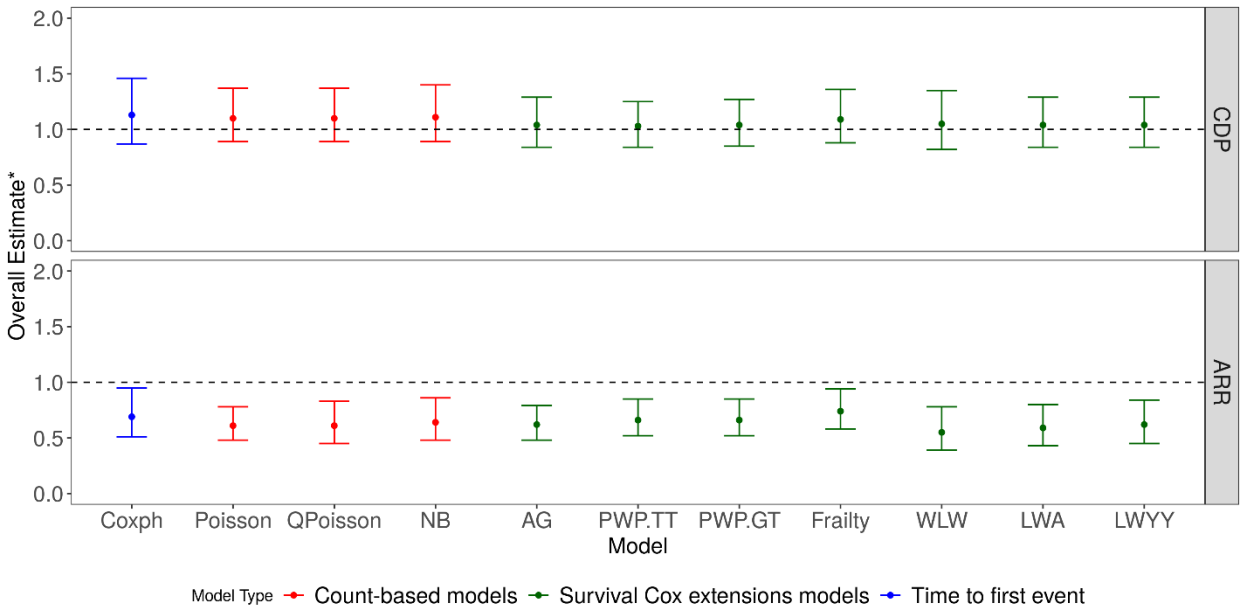
3.3.3.1. Overall effect of disease course on CDP and ARR

In figure 4, is presented the estimates and its corresponding 95% confidence intervals for CDP and ARR endpoints by disease course for reviewed models and the CoxPH.

From a qualitative point of view, the results of the model appeared to be consistent. CDP rates tend to be similar in both groups, while the annualized relapse rate was lower in participants with a diagnosed SPMS. When comparing time to first event methods with recurrent event methods, CoxPH model had wider confidence intervals. Furthermore, in survival cox extensions, it was observed that estimated HR or RR tended to be similar in magnitude within conditionals and within marginal models, except for frailty models. Count-based models showed similar results for both endpoints, apart from the Poisson model which presented smaller confidence intervals in the ARR.

Even though the recurrent CDP were not significantly different in both groups, the number of relapses was significantly lower in the SPMS group in all fitted models (Figure 4). The risk of relapses was 43% (WLW), 39% (LWA, Poisson and Quasi-Poisson), 38% (Andersen-Gill and LWYY), 36% (Negative Binominal), 34% (PWP-TT and PWP-GT) and 26% (Frailty) lower in the SPMS group compared with the RRMS group (Table S10).

On further details, when individuals were assumed to be simultaneously at risk for all events, WLW and LWA resulted in a 43% (HR = 0.57; 95% CI: 0.39 - 0.78) and 39% (HR = 0.61; 95% CI: 0.44 - 0.82) lower risk of relapse in the SPMS group than in the RRMS, respectively. By conditioning by covariates, the AG model revealed that patients with SPMS diagnosis had a reduction of 38% on the risk of relapses compared with the RRMS group (HR = 0.62; 95% CI: 0.48 - 0.79). Both PWP-TT and PWP-GT models estimated a 34% reduction in the risk of relapse (HR = 0.66, 95% CI: 0.52 - 0.85) in SPMS participants when an event-specific baseline hazard was assumed. This risk was 26% lower in the Frailty model when accounting for unmeasured heterogeneity (HR = 0.74; 95% CI: 0.58 - 0.94). The Poisson regression yielded a decreasing risk of 39% (RR = 0.61; 95% CI: 0.48 - 0.78). By correcting overdispersion, the Quasi-Poisson model yielded the same estimate rate ratio as Poisson model but with different confidence intervals (RR = 0.61; 95% CI: 0.45 - 0.83). The risk of relapses was also reduced in the SPMS group in the NB model (RR = 0.64; 95% CI 0.48 - 0.86) while correcting for heterogeneity and assuming gamma distribution. Finally, when not considering the dependency structure, the LWYY model estimated a similar risk reduction of 38% (RR: 0.62; 95% CI: 0.45 - 0.84) in the SPMS group compared to RRMS than the AG model.



Abbreviations: Cox Proportional Hazard (CoxPH), Quasi-Poisson (QPoisson), Negative Binomial model (NB), Andersen Gill (AG), Prentice-Williams-Peterson (PWP-GT), Prentice-Williams-Peterson (PWP-TT), Wei-Lin-Weissfeld (WLW), Lee-Wei-Amato (LWA), Lin-Wei-Yang-Ying (LWYY). *Hazard ratio (HR) for CoxPH, AG, PWP-TT, PWP-GT, Frailty and WLW; Rate Ratio (RR) for LWYY, NB, Poisson, and Quasi-Poisson.

Figure 4. Overall estimate disease course effect on confirmed disability progression (CDP) and Annualized relapse rate (ARR) adjusting by covariates.

3.3.3.2. Event-specific estimates of CDP and ARR

Given that PWP, WLW and the PRBC models are stratified by event number, these models are adequate to obtain the event-specific estimates of the regression parameters. Only the effects for the first 3 events were reported (Table 5).

Event-specific estimates of the first event are the same in the presented models, which are also the same estimates that are obtained with the CoxPH (Table S2). Similarly to the AG and LWYY in overall estimates, the PWP and PCRB models provide same point estimates with few differences in the 95% confidence intervals.

The risk of second and third CDP was not significant different in both groups for all models. In contrast, the risk of the second and third relapse was 49% and 65% significantly lower in SPMS participants than in RRMS group for the WLW model. In the PWP and PRBC models, there was a significant decrease in the risk of relapse by 35% for the second event in the SPMS group.

Table 8. Overall and event-specific effect on confirmed disability progression (CDP) and Annualized relapse rate (ARR) adjusting by covariates.

		CDP			ARR		
Model	**Event**	**Estimate[1]**	**95% CI**	**P value**	**Estimate**	**95% CI**	**P value**
PWP-TT	Overall	1.03	0.84 - 1.25	0.7871	0.66	0.52 - 0.85	0.0013
	1	1.12	0.87 - 1.45	0.374	0.69	0.51 - 0.95	0.0206
	2	0.88	0.56 - 1.39	0.5945	0.65	0.40 - 1.04	0.0702
	3	0.62	0.21 - 1.84	0.3871	0.75	0.33 - 1.69	0.489
WLW	Overall	1.06	0.82 - 1.35	0.7056	0.57	0.39 - 0.78	0.0007
	1	1.14	0.90 - 1.46	0.2722	0.69	0.50 - 0.97	0.0316
	2	0.92	0.60 - 1.42	0.7054	0.51	0.29 - 0.90	0.0205
	3	0.64	0.26 - 1.57	0.3323	0.35	0.14 - 0.87	0.0246
PRCB	Overall	1.03	0.84 - 1.25	0.7871	0.66	0.52 - 0.85	0.0013
	1	1.12	0.88 - 1.44	0.3578	0.69	0.50 - 0.97	0.0316
	2	0.88	0.59 - 1.33	0.5577	0.65	0.41 - 1.00	0.0525
	3	0.62	0.23 - 1.65	0.3353	0.75	0.36 - 1.57	0.4439

Abbreviations: Prentice-Williams-Peterson Total Time (PWP-TT), Wei-Lin-Weissfeld (WLW), Partially conditional rate-based (PCRB). [1]Hazard ratio (HR) for PWP-TT and WLW and Rate Ratio (RR) for PRCB

4. DISCUSSION

4.1. Literature review findings

This study has identified and described the main statistical models in recurrent events through a comprehensive literature review. It has been seen that several publications that have addressed the application of various recurrent event methods, as well as highlighted the problems faced by researchers when applying the methods in epidemiological and clinical settings (4,6,8,18–20,24). It was also observed that some studies continues modelling with time-to-first event or other mentioned traditional methodologies not recommended for recurrent event data (61–63). In addition to the previous documented reasons, it can also be suggested that the relative lack of published guidelines or recommendations on appropriate methods for analyzing such events may be limiting the use of recurrent event methods. To date, only one guideline issued by EMA has recalled the general use of recurrent event methods in several chronic diseases (31).

Although most reviewed articles followed the general classification of conditionals and marginal models, it was noticed a lack of agreement when classifying recurrent event methods in lower

hierarchical levels (8,19,22,26). For instance, some original research articles have not mentioned the specific name of the employed model (i.e., LWYY model), but instead only mention a higher hierarchy of the model (i.e. marginal rate model) (9,13,19,62). Moreover, it was found that several articles did not justify the use of the selected model (9,57). As this may affect the understanding and future application of this methods for health professional researchers, a simplified classification has been proposed in this study.

Furthermore, the literature review enabled to identify several statistical models (8,23,24,59). As expected, the AG model was the most documented model. It is one of the earliest described in recurrent events methods as well as well-known for its flexibility and relatively easy applicability (4,6,8,22,23). Other conditional cox's extensions, such as the PWP and Frailty models, have been found to be common in observational designs, in which the PWP also allows the estimation of the separate risk for each event while the frailty properly account for the dependency within subjects and unobserved heterogeneity among patients (6,8,18,26). Although the marginal WLW was one of the most cited models in the literature on recurrent events, it is of particular interest that most articles were focused on the problems of its application to successive events, rather than its advantages (6,8,18). Most marginal and count-based models were less mentioned. However, it was observed that the NB, LWYY and PCRB models are gaining popularity in experimental designs during the last decade especially given their ability to conserve randomizations benefits (21,24,34). The Poisson regression is less documented probably due to its strict assumptions and overdispersion problems, but its variance-corrected versions (Quasi-Poisson and GEE model) are often implemented in recurrent events (4,6,8,54).

4.2. Model comparison

Identified statistical models were applied to the MSOAC clinical database for comparing estimates, provide interpretation examples and present their respective advantages and disadvantages. Two main MS endpoints were employed to provide additional interpretation and a detailed explanation of data structuring was conducted before fitting reviewed models (6,8,42).

As it was previously documented in other studies, the common and event-specific estimates for CDP and ARR endpoints have been shown to be relatively consistent across models and even when comparing time-to first event (CoxPH) with recurrent event methods (6,8,20,22,26,46,64). However, it was seen that recurrent event models presented smaller 95% CI, as these improve statistical precision of the estimates by considering all the available event data for effect estimation (4,8,65). It was also observed that event-specific estimates of the PWP, WLW and PCRB

produced identical estimates of the effect for the first event than the CoxPH model. This is explained by the fact that they use the same definition of risk set for the first event (6,8,42).

Given its simplicity of application and interpretation, the Poisson regression is often used in recurrent events analyses. As it was found in this study, smaller confidence intervals in the effect measure can be obtained when it is compared with the CoxPH (4,6). However, estimates may be biased and overestimated due to the overdispersion of the data, as is the case with relapses (overdispersion ratio = 1.19) (6). For that reason, Poisson regression is not recommended in analyses in over dispersed recurrent event data or if the Poisson distribution assumption is not met (4,6). Quasi-Poisson and GEE Poisson has proven useful in MS and other recurrent event studies despite the fact that these models do not use event timing information (4,19,24,66). In the presented analyses, overdispersion correction in the Quasi-Poisson model can be observed in the different confidence intervals estimated in the relapse rates.

Without overdispersion, the negative binomial model coincides with the Poisson model (8,46). However, the NB has less stronger assumptions and appears more plausible as assumes that each individual has their own event rate while accounting for within-subject correlation (46). These characteristics may explain the estimate differences obtained in this study between the NB with the other count-based models. Compared to classical conditional models in recurrent events (i.e. AG model), the NB regression with random effects requires a much simpler data structure, is easier to implement, and usually gives comparable performance for assessing the overall treatment effects (6). It has been claimed as an accurate model for recurrent event analyses in MS and its use in RTCs has increased in the last decade (6,21,67,68).

The AG model is a flexible option which makes the full use of the data (6,19,45). The overall estimate provided by the AG model was considerable different to the CoxPH model even without considering the use of the robust variance for improving the estimate precision (6). Although this model is particularly criticized due to the assumption of common baseline hazards for all events, it was satisfactory fulfilled in this study for both outcomes. By considering relevant covariates, the AG model has shown a broader application in health research for the estimation of the overall effect, particularly when there is no clear biological mechanism underlying the relation between the first and subsequent events (6,19,30,49).

In contrast to the AG model, the PWP accounts for potential strong dependence between events when adding the stratification term (8,18). In the context of MS, it is of interest when the research question is based on estimating the effect in patients with more than one CDP or relapses or when

the interest rely in estimating the separate risk of each event (6,18). Even though the PWP-TT and PWP-GT models presented similar estimates, the interpretation differs due to different time scales used to fit the model. The renewal approach of the PWP-GT makes its interpretation unreliable for MS and other chronic diseases (8). Moreover, the event-specific estimates for the PWP are more recommended when the risk of CDP and relapses varies between events and when there are few recurrent events per subject (19). The latter condition may not apply to the MSOAC database with individuals with more than three CDP or relapses.

Frailty models are suitable when some participants are intrinsically more susceptible to experiencing recurrent events than others, as account for within-subject correlation between events and unobserved heterogeneity among patients by adding a random covariate (18,19,41,49). A frailty variance close to zero implies low correlation between the event times (18). In this study, the frailty variance for CDP and relapses was 0.00004 and 0.00003, respectively which means that the application of frailty models may be not justified. Therefore, the differences observed between the frailty model estimates with other models can be explained by model's assumptions and not by the presence of heterogeneity (6,49).

As mentioned in previous studies, the WLW approach overestimated the measured effect both in common and event-specific estimate analyses (4,6,8,26). It can be explained by the strong assumption that all patients are included in the risk set for each CDP and relapse stratum, allowing the effect on earlier events to affect subsequent events (4,6,52). The same applies to the LWA model, which may also provide biased estimates effects (4,8,26)**.** Although several authors argue that WLW and LWA are not suitable for recurrent event analyses in most health settings, these models should be considered for the analysis of unordered events and clustered data (4,26,53).

The marginals LWYY and PCRB have provided the same estimates as the conditionals AG an PWP because no time-dependent covariates were considered in the conditional model (8,19,59). Moreover, even though the LWYY often yield similar effect estimates to the NB model, the results were not the same for both endpoints probably due to differences in the event rate function (21). Despite similarities with conditional methods, marginals mean/rate models tend to be preferred as the effect measure is more interpretable than the hazard and have less stronger assumptions with an arbitrary dependence structure between recurrent events (8). The PCRB can be considered a suitable complementary analysis to obtain a comprehensive overview of the overall effect and to study the effects of the exposure variable on time to subsequent events (8).

4.3. Factors influencing the model choice

Comparison of models based on theory and estimates results does not allow to choose a model by itself. Several other factors should be considered before selecting a model for recurrent events data. As presented in this study, the application of mathematical criteria may allow to get extra information about the model fitness (20,41). However, it is already known that models with event-specific baseline hazard usually present a better fit (20,42).

Another key factor for model choice is the type of study design (22). RCTs are often interested in models based on total time scale (i.e. NB, LWYY and PCRB). Conversely, in prospective observational studies, the interest may lie in developing conditional models (i.e. AG and PWP) to better understand factors associated to the event and the correct specification of the recurrent event process (8). Frailty models can be recommended to mitigate the risk of selection bias when individuals with higher risks of recurrent events may contribute disproportionately to the analysis due to previous events (69). Moreover, it is important to consider the simulation or prediction purposes, as fully specified models with fixed and time-varying covariates are often preferred (22).

Completeness and type of data should be also taken into consideration at the moment of applying recurrent event models due to data with missing components often require more assumptions for their analysis (22,30). In addition, it is essential to understand the dependence structure between recurrent events and the biological processes underlying the recurrent events (19,28,30,32). The set of covariates and the number of events available need to be also considered and particularly in conditional methods (18,30). Result interpretation and availability of commercially accessible statistical software may be also relevant because not all statistical tests were explicitly developed for recurrent event analyses (33,49).

4.4. Strengths and limitations of the study

As far it is known, this study is the first well-structured literature review that has been conducted for recurrent events with a subsequent application of main statistical models on real data. It has allowed not only to identify methods from a broad perspective of articles, including theses, conferences, and original research articles, but also to provide a detailed description and practical application of each recurrent event model, offering interpretation resources for epidemiologists and other health professionals. Its application to the MSOAC database allowed the generation of more accurate results because of its large number of participants and due to the high-quality outcomes measures in the context of regulated clinical trials. Additionally, unlike most reviewed studies focusing solely on relapses or disability progression, this research applies methods to both main endpoints of MS, closely mimicking general practices in MS research.

However, the study presents some limitations. The literature review process, may be subject to bias due to lack of independent revisors in data collection and extraction. Another limitation is the absence of a simulation study to further investigate the properties of each evaluated method. This would have provided more information about robustness and performance in various hypothetical scenarios, improving the applicability of the results. Although this study was not intended to generalize the results of the model but rather to perform the analyses under ideal conditions, several limitations related to the application phase can be mentioned, such as the use of placebo arms, non-informative censoring, unmeasured confounders, and selection bias due to the selected participants.

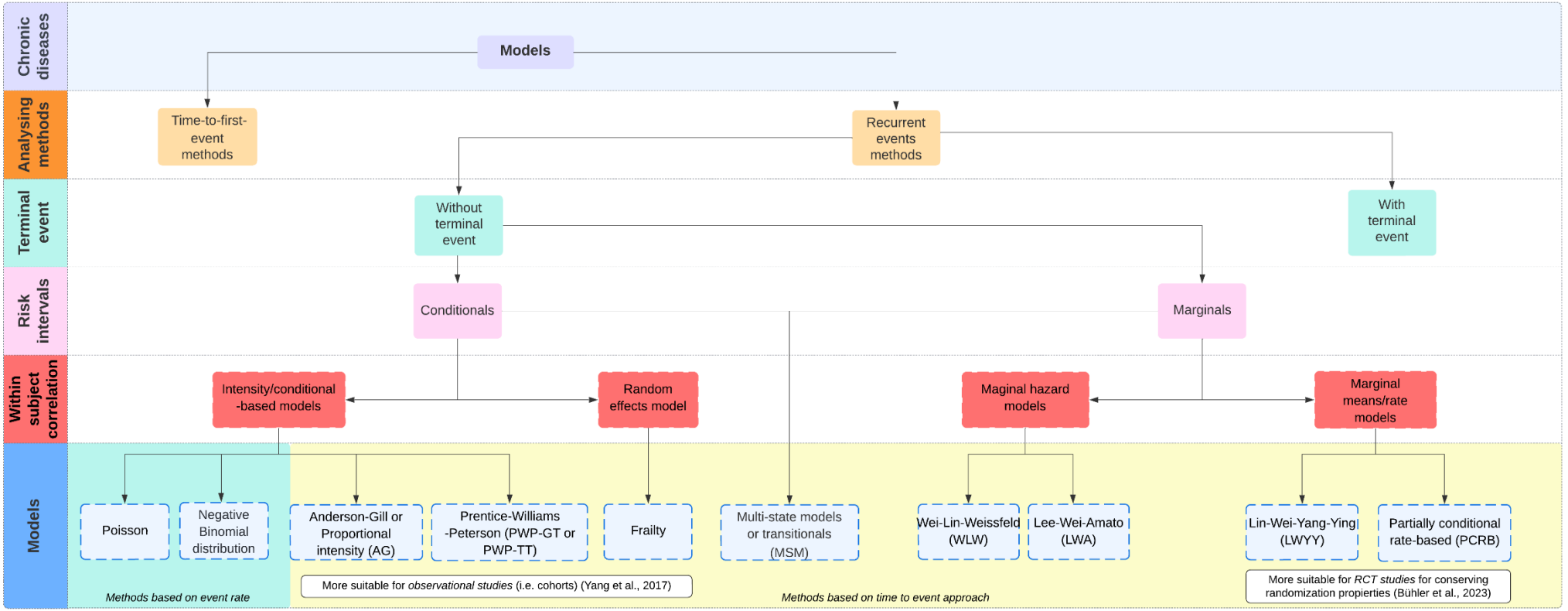
5. CONCLUSION

This study overviewed the main statistical models of recurrent events and provides classification, interpretation, and methodological guidance. By considering all CDP and relapses events, recurrent events methods can help to gain insights into the MS disease process compared with the traditional survival and count-based methods. No single model is indicated to address most of MS research questions as all models have their own assumptions and characteristics. Its choice may greatly vary depending on the research study and other factors.

However, it is possible to suggest some models based on general advantages over other models. The NB model or the marginal LWYY can be recommended for its application in both CDP and relapses in experimental studies over the AG and Poisson models. The PWP-TT or PCRB models, can be more accurate than the marginals WLW or LWA if the interest relies in measuring the overall effect in patients with one or more events or even when event-specific estimates are required. In addition, the frailty model may be indicated particularly in observational studies when heterogeneity in patients may affect the effect estimation above other conditional models.

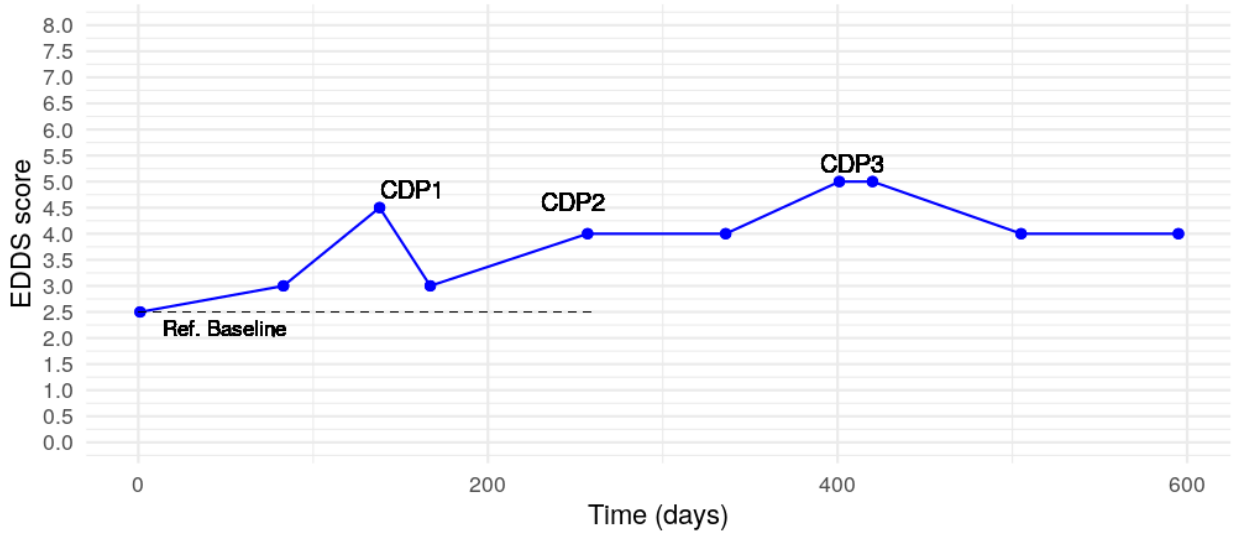
To conclude, some recommendations can be done: i) researchers should give priority to recurrent event methods when developing the statistical plan to avoid the use of inefficient methods and make better use of data. Certain recurrent event methods can be also used to study the time until the first event; ii) the development of further official guidelines are required. This may promote better practices in the scientific community and consequently improve result interpretation and reproducibility research; iii) research articles with recurrent event methods should clearly specify and justify the selected model based on epidemiological aspects or the research question addressed and not only in its statistical benefits; iv) consider the data structure not only when fitting models that requires specific layouts, but also when comparing model fitness; v) use up-to date commands on available software such as R.

Supplementary materials



Source: Author creation with Lucid®

Figure S5. Flowchart with main characteristics of recurrent events methods and main models.



Source: Author creation with Rstudio®

*CDP: confirmed disability progression.

Figure S6. Derivation of the disability progression counting for recurrent events since reference baseline.

Table S9. R commands for recurrent events implementation

Model	R command
CoxPH	coxph(Surv(TSTART, TSTOP, EVENT) ~ (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION), data = subset(data_layout_2, SEVENT == 1))
Poisson	glm(COUNT ~ offset(log(LENGHT.TIME)) + DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION, family = poisson(link = "log"), data = data_layout_1)
Quasi-Poisson	glm(COUNT ~ offset(log(LENGHT.TIME)) + DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION, family = quasipoisson(link = "log"), data = data_layout_1)
NB	glm.nb(COUNT ~ offset(log(LENGHT.TIME)) + DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION, data = data_layout_1)
AG	coxph(Surv(TSTART, TSTOP, EVENT) ~ (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION), data = data_layout_2)
PWP-TT	coxph(Surv(TSTART, TSTOP, EVENT) ~ (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION) + strata(SEVENT), data = data_layout_2)
PWP-GT	*coxph(Surv(TGAP, EVENT) ~ (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION) + strata(SEVENT), data = data_layout_2)*
PWP-TT event specific estimates	*coxph(Surv(TSTART, TSTOP, EVENT) ~ strata(SEVENT) / (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION), data = subset(data_layout_2, SEVENT <= 3))*
Frailty	*coxph(Surv(TSTOP, EVENT) ~ (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION) + frailty(USUBJID), data = data_layout_2)*
WLW	coxph(Surv(TSTOP, EVENT) ~ (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION) + cluster(USUBJID) + strata(SEVENT), data = data_layout_3)
WLW event specific estimates	coxph(Surv(TSTOP, EVENT) ~ strata(SEVENT)/(DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION) + cluster(USUBJID), data = subset(data_layout_3, SEVENT <= 3)
LWA	coxph(Surv(TSTOP, EVENT) ~ (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION) + cluster(USUBJID), data = data_layout_3)
LWYY	coxph(Surv(TSTART, TSTOP, EVENT) ~ (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION) + cluster(USUBJID), data = data_layout_2)
PCRB event-specific estimates	coxph(Surv(TSTART, TSTOP, EVENT) ~ strata(SEVENT) / (DISEASE_COURSE + AGE + SEX + RACE + DISEASE_DURATION) + cluster(USUBJID), data = *subset(data_layout_2, SEVENT <= 3))*

Abbreviations: Cox Proportional Hazard (CoxPH), Negative Binomial model (NB), Andersen Gill (AG), Prentice-Williams-Peterson (PWP-GT), Prentice-Williams-Peterson (PWP-TT), Wei-Lin-Weissfeld (WLW), Lee-Wei-Amato (LWA), Lin-Wei-Yang-Ying (LWYY), Partially conditional rate-based (PCRB)

Table S10. Overall effect on disability progression and relapse.

	CDP			ARR		
Model	**Estimate[1]**	**95% CI**	**P value**	**Estimate**	**95% CI**	**P value**
CoxPH	1.12	0.87 - 1.46	0.3576	0.69	0.51 - 0.95	0.0206
Poisson	1.10	0.89 - 1.37	0.3641	0.61	0.48 - 0.78	0.00008
QPoisson	1.10	0.89 - 1.37	0.3740	0.61	0.45 - 0.83	0.0016
NB	1.11	0.89 - 1.40	0.3539	0.64	0.48 - 0.86	0.0032
AG	1.04	0.84 - 1.29	0.7153	0.62	0.48 - 0.79	0.0001
PWP.TT	1.03	0.84 - 1.25	0.7871	0.66	0.52 - 0.85	0.0013
PWP.GT	1.04	0.85 - 1.27	0.7030	0.66	0.52 - 0.85	0.0010
Frailty	1.09	0.88 - 1.36	0.425	0.74	0.58 - 0.94	0.0147
WLW	1.06	0.82 - 1.35	0.7056	0.57	0.39 - 0.78	0.0007
LWA	1.05	0.84 - 1.29	0.6985	0.61	0.44 - 0.82	0.0002
LWYY	1.04	0.84 - 1.31	0.675	0.62	0.45 - 0.84	0.0021

Abbreviations: Cox Proportional Hazard (CoxPH), Quasi-Poisson (QPoisson), Negative Binomial model (NB), Andersen Gill (AG), Prentice-Williams-Peterson Gap Time (PWP-GT), Prentice-Williams-Peterson Total Time (PWP-TT), Wei-Lin-Weissfeld (WLW), Lee-Wei-Amato (LWA), Lin-Wei-Yang-Ying (LWYY).

[1] Hazard ratio (HR) for CoxPH, AG, PWP-TT, PWP-GT, Frailty and WLW; Rate Ratio for LWYY and NB; IRR for poisson and QPoisson.

Table S11. Model comparison using mathematical indices.

			CDP			ARR		
Model	**Data layout**	**Event**	**LL**	**AIC**	**BIC**	**LL**	**AIC**	**BIC**
Poisson	Layout 1	Overall	-1174.3	2360.6	2391.6	-1223.3	2458.5	2489.5
QPoisson		Overall	-1174.3	2270.7	-	-1223.2	1581.5	-
NB		Overall	-1170.3	2354.6	2390.7	-1172.4	2358.7	2394.8
CoxPH	Layout 2	1st event	-3106.4	6222.9	6243.6	-2665.6	5341.3	5361.2
AG		Overall	-4520.8	9051.5	9074.1	-4316.2	8642.5	8664.8
PWP.TT		Overall	-4040.8	8091.7	8114.2	-3687.6	7385.2	7407.6
PWP.GT		Overall	-4089.3	8188.6	8211.1	-3766.4	7542.9	7565.2
Frailty		Overall	-4719.8	9449.6	9472.2	-4505.2	9020.4	9042.8
LWYY		Overall	-4520.8	9051.5	9074.1	-4316.2	8642.5	8664.8
PRCB.SP		Event-specific	-4004.7	8023.3	8054.8	-3615.6	7245.3	7276.1
PWP.SP		Event-specific	-4004.7	8023.3	8054.8	-3615.6	7245.3	7276.1
WLW	Layout 3	Overall	-4567.0	9144.1	9166.8	-4322.0	8654.0	8676.5
LWA		Overall	-6039.8	12089.6	12112.4	-5921.1	11852.3	11874.7
WLW.SP		Event-specific	-4454.9	8923.7	8955.4	-4101.2	8216.4	8247.4

Abbreviations: Log-likelihood (LL), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC)

Due to the database's layout differ in the total number of rows, the model comparison criteria can be only conducted between models that were fitted with the same type of database. Data layout for count based models have found the lowest AIC for Quasi-Poisson model for both endpoints. However, the NB model had the best fitted model when comparing by log-likelihood and BIC criteria. On the other hand, without considering the time to first event CoxPH model, recurrent event data layout has shown that the PWP-TT model was had the lowest value for all criteria for both endpoints in overall and event-specific types of models. Lastly, the WLW had the smallest values for the three indices when data layout 3 was used.

References

1. MS International Federation. MS International Federation. 2020 [cited 2024 Mar 8]. Atlas of MS 2020 - Epidemiology report. Available from: https://www.msif.org/resource/atlas-of-ms-2020/

2. Wallin MT, Culpepper WJ, Nichols E, Bhutta ZA, Gebrehiwot TT, Hay SI, et al. Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019 Mar 1;18(3):269–85.

3. Kamel FO. Factors Involved in Relapse of Multiple Sclerosis. J Microsc Ultrastruct. 2019 Sep;7(3):103.

4. Mieno MN, Yamaguchi T, Ohashi Y. Alternative statistical methods for estimating efficacy of interferon beta-1b for multiple sclerosis clinical trials. BMC Med Res Methodol. 2011 May 26;11(1):80.

5. Hernandez L, O'Donnell M, Postma M. Predictors of Health Utility in Relapsing–Remitting and Secondary-Progressive Multiple Sclerosis: Implications for Future Economic Models of Disease-Modifying Therapies. Pharmacoeconomics. 2021;39(2):243–56.

6. Wang YC, Meyerson L, Tang YQ, Qian N. Statistical methods for the analysis of relapse data in MS clinical trials. J Neurol Sci. 2009 Oct 15;285(1):206–11.

7. National Multiple Sclerosis Society. National Multiple Sclerosis Society. 2024 [cited 2024 Mar 4]. Types of Multiple Sclerosis. Available from: https://www.nationalmssociety.org/What-is-MS/Types-of-MS

8. Bühler A. Comparison of Time-to-First-Event and Recurrent Event Methods in Multiple Sclerosis Trials [Internet]. arXiv; 2019 [cited 2024 Jan 17]. Available from: http://arxiv.org/abs/2111.01937

9. Lorscheider J, Jokubaitis VG, Spelman T, Izquierdo G, Lugaresi A, Havrdova E, et al. Anti-inflammatory disease-modifying treatment and short-term disability progression in SPMS. Neurology. 2017 Sep 5;89(10):1050–9.

10. Reeve K, On BI, Havla J, Burns J, Gosteli-Peter MA, Alabsawi A, et al. Prognostic models for predicting clinical disease progression, worsening and activity in people with multiple sclerosis.

Cochrane Database Syst Rev [Internet]. 2023 [cited 2024 Mar 18];(9). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013606.pub2/full

11. Kalincik T, Cutter G, Spelman T, Jokubaitis V, Havrdova E, Horakova D, et al. Defining reliable disability outcomes in multiple sclerosis. Brain. 2015 Nov 1;138(11):3287–98.

12. Marrie RA, Sormani MP, Apap Mangion S, Bovis F, Cheung WY, Cutter GR, et al. Improving the efficiency of clinical trials in multiple sclerosis. Mult Scler J. 2023 Aug 1;29(9):1136–48.

13. Naegelin Y, Naegelin P, von Felten S, Lorscheider J, Sonder J, Uitdehaag BMJ, et al. Association of Rituximab Treatment With Disability Progression Among Patients With Secondary Progressive Multiple Sclerosis. JAMA Neurol. 2019 Mar 1;76(3):274–81.

14. Healy BC, Glanz BI, Swallow E, Signorovitch J, Hagan K, Silva D, et al. Confirmed disability progression provides limited predictive information regarding future disease progression in multiple sclerosis. Mult Scler J - Exp Transl Clin. 2021 Apr 11;7(2):2055217321999070.

15. Healy BC, Engler D, Glanz B, Musallam A, Chitnis T. Assessment of Definitions of Sustained Disease Progression in Relapsing-Remitting Multiple Sclerosis. Mult Scler Int. 2013;2013:189624.

16. Lavery AM, Verhey LH, Waldman AT. Outcome Measures in Relapsing-Remitting Multiple Sclerosis: Capturing Disability and Disease Progression in Clinical Trials. Mult Scler Int. 2014 May 4;2014:e262350.

17. Twisk J, Smidt N, de Vente W. Applied analysis of recurrent events: a practical overview. J Epidemiol Community Health. 2005 Aug;59(8):706–10.

18. Thenmozhi M, Jeyaseelan V, Jeyaseelan L, Isaac R, Vedantam R. Survival analysis in longitudinal studies for recurrent events: Applications and challenges. Clin Epidemiol Glob Health. 2019 Jun 1;7(2):253–60.

19. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. Int J Epidemiol. 2015 Feb;44(1):324–33.

20. Yadav CP, Lodha R, Kabra SK, Sreenivas V, Sinha A, Khan MA, et al. Comparison of statistical methods for recurrent event analysis using pediatrics asthma data. Pharm Stat. 2020 Nov;19(6):803–13.

21. Bühler A, Wolbers M, Model F, Wang Q, Belachew S, Manfrini M, et al. Recurrent disability progression endpoints in multiple sclerosis clinical trials. Mult Scler Houndmills Basingstoke Engl. 2023 Jan;29(1):130–9.

22. Cook RJ, Lawless J. The Statistical Analysis of Recurrent Events [Internet]. New York, NY: Springer; 2007 [cited 2024 Feb 13]. (Statistics for Biology and Health). Available from: http://link.springer.com/10.1007/978-0-387-69810-6

23. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. Ann Stat. 1982;10(4):1100–20.

24. Rogers JK, Pocock SJ, McMurray JJV, Granger CB, Michelson EL, Östergren J, et al. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. Eur J Heart Fail. 2014 Jan;16(1):33–40.

25. Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. Biometrics. 2007 Mar;63(1):78–87.

26. Kelly PJ, Lim LL. Survival analysis for recurrent event data: an application to childhood infectious diseases. Stat Med. 2000 Jan 15;19(1):13–33.

27. Rogers DJ. The Analysis of Recurrent Events: A Summary of Methodology. 2016;

28. Charles-Nelson A, Katsahian S, Schramm C. How to analyze and interpret recurrent events data in the presence of a terminal event: An application on readmission after colorectal cancer surgery - Charles-Nelson - 2019 - Statistics in Medicine - Wiley Online Library. 2019 [cited 2024 Feb 13]; Available from: https://onlinelibrary.wiley.com/doi/10.1002/sim.8168

29. Kim S, Schaubel DE, McCullough KP. A C-index for recurrent event data: Application to hospitalizations among dialysis patients. Biometrics. 2018 Jun;74(2):734–43.

30. Rauch G, Kieser M, Binder H, Bayes-Genis A, Jahn-Eimermacher A. Time-to-first-event versus recurrent-event analysis: points to consider for selecting a meaningful analysis strategy in

clinical trials with composite endpoints. Clin Res Cardiol Off J Ger Card Soc. 2018 May;107(5):437–43.

31. Committee for Medicinal Products for Human Use (CHMP. Qualification opinion of clinically interpretable treatment effect measures based on recurrent event endpoints that allow for efficient statistical analyses. 2020;(European Medicines Agengy (EMA). Committee for Medicinal Products for Human Use (CHMP);).

32. Claggett B, Pocock S, Wei LJ, Pfeffer MA, McMurray JJV, Solomon SD. Comparison of Time-to-First Event and Recurrent-Event Methods in Randomized Clinical Trials. Circulation. 2018 Aug 7;138(6):570–7.

33. Anker SD, McMurray JJV. Time to move on from 'time-to-first': should all events be included in the analysis of clinical trials? Eur Heart J. 2012 Nov 1;33(22):2764–5.

34. Akacha M, Binkowitz B, Claggett B, Hung HMJ, Mueller-Velten G, Stockbridge N. Assessing Treatment Effects That Capture Disease Burden in Serious Chronic Diseases. Ther Innov Regul Sci. 2019 May;53(3):387–97.

35. Seker BIO, Reeve K, Havla J, Burns J, Gosteli MA, Lutterotti A, et al. Prognostic models for predicting clinical disease progression, worsening and activity in people with multiple sclerosis. Cochrane Database Syst Rev [Internet]. 2020 [cited 2024 Mar 18];2020(5). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7389149/

36. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023) [Internet]. 2023 [cited 2024 Apr 25]. Available from: www.training.cochrane.org/handbook.

37. Gani R, Nixon RM, Hughes S, Jackson CH. Estimating the rates of disability progression in people with active relapsing-remitting multiple sclerosis. J Med Econ. 2007 Jan 1;10(2):79–89.

38. Critical Path Institute. Multiple Sclerosis Outcome Assessments Consortium (MSOAC) [Internet]. 2024 [cited 2024 Feb 21]. Available from: https://c-path.org/multiple-sclerosis-outcome-assessments-consortium-msoac-placebo-database-faq/

39. Kalinowski A, Cutter G, Bozinov N, Hinman JA, Hittle M, Motl R, et al. The timed 25-foot walk in a large cohort of multiple sclerosis patients. Mult Scler J. 2022 Feb 1;28(2):289–99.

40. Ramagopalan SV, Dobson R, Meier UC, Giovannoni G. Multiple sclerosis: risk factors, prodromes, and potential causal pathways. Lancet Neurol. 2010 Jul;9(7):727–39.

41. Ullah S, Gabbett TJ, Finch CF. Statistical modelling for recurrent events: an application to sports injuries. Br J Sports Med. 2014 Sep;48(17):1287–93.

42. Sousa-Ferreira I, Abreu AM. A review of Cox' s model extensions for multiple events. In 2019 [cited 2024 Apr 12]. Available from: https://www.semanticscholar.org/paper/A-review-of-Cox%27-s-model-extensions-for-multiple-Sousa-Ferreira-Abreu/2ab71f1438135a4a9fb3accfd3b3dab28cb74cf2

43. Yang W, Jepson C, Xie D, Roy JA, Shou H, Hsu JY, et al. Statistical Methods for Recurrent Event Analysis in Cohort Studies of CKD. Clin J Am Soc Nephrol. 2017 Dec;12(12):2066.

44. Rondeau V, Filleul L, Joly P. Nested frailty models using maximum penalized likelihood estimation. Stat Med. 2006 Dec 15;25(23):4036–52.

45. Yang X, Lu X, Li W, Huang Y, Huang X, Lu M, et al. Endovascular treatment for symptomatic stent failures in long-segment chronic total occlusion of femoropopliteal arteries. J Vasc Surg. 2014 Aug;60(2):362–8.

46. Keene. Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited - PubMed. 2008 [cited 2024 Feb 9]; Available from: https://pubmed.ncbi.nlm.nih.gov/18591336/

47. Sormani MP, Bruzzi P, Miller DH, Gasperini C, Barkhof F, Filippi M. Modelling MRI enhancing lesion counts in multiple sclerosis using a negative binomial model: implications for clinical trials. J Neurol Sci. 1999 Feb 1;163(1):74–80.

48. Tang Y, Fitzpatrick R. Sample size calculation for the Andersen-Gill model comparing rates of recurrent events. Stat Med. 2019 Oct 30;38(24):4819–27.

49. Guo Z, Gill TM, Allore HG. Modeling repeated time-to-event health conditions with discontinuous risk intervals: an example of a longitudinal study of functional disability among older persons. Methods Inf Med. 2008;47(2):107–16.

50. Lintu MK, Kamath A. Performance of recurrent event models on defect proneness data. Ann Oper Res. 2022 Aug 1;315(2):2209–18.

51. Patson N, Mukaka M, Kazembe L, Eijkemans MJC, Mathanga D, Laufer MK, et al. Comparison of statistical methods for the analysis of recurrent adverse events in the presence of non-proportional hazards and unobserved heterogeneity: a simulation study. BMC Med Res Methodol. 2022 Jan 20;22(1):24.

52. Metcalfe C, Thompson SG. Wei, Lin and Weissfeld's marginal analysis of multivariate failure time data: should it be applied to a recurrent events outcome? Stat Methods Med Res. 2007 Apr;16(2):103–22.

53. Tikkanen MJ, Szarek M, Fayyad R, Holme I, Cater NB, Faergeman O, et al. Total Cardiovascular Disease Burden: Comparing Intensive With Moderate Statin Therapy: Insights From the IDEAL (Incremental Decrease in End Points Through Aggressive Lipid Lowering) Trial. J Am Coll Cardiol. 2009 Dec 15;54(25):2353–7.

54. Newsome SD, Mokliatchouk O, Castrillo-Viguera C, Naylor ML. Matching-adjusted comparisons demonstrate better clinical outcomes in patients with relapsing multiple sclerosis treated with peginterferon beta-1a than with teriflunomide. Mult Scler Relat Disord. 2020 May;40:101954.

55. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, et al. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. N Engl J Med. 2017 Jan 19;376(3):221–34.

56. Rassouli F, Baty F, Stolz D, Albrich WC, Tamm M, Widmer S, et al. Longitudinal change of COPD assessment test (CAT) in a telehealthcare cohort is associated with exacerbation risk. Int J Chron Obstruct Pulmon Dis. 2017;12:3103–9.

57. Lizak N, Malpas CB, Sharmin S, Havrdova EK, Horakova D, Izquierdo G, et al. Association of Sustained Immunotherapy With Disability Outcomes in Patients With Active Secondary Progressive Multiple Sclerosis. JAMA Neurol. 2020 Nov 1;77(11):1398–407.

58. Mogensen UM, Gong J, Jhund PS, Shen L, Køber L, Desai AS, et al. Effect of sacubitril/valsartan on recurrent events in the Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart Failure trial (PARADIGM-HF). Eur J Heart Fail. 2018 Apr;20(4):760–8.

59. Furberg JK, Rasmussen S, Andersen PK, Ravn H. Methodological challenges in the analysis of recurrent events for randomised controlled trials with application to cardiovascular events in LEADER. Pharm Stat. 2022;21(1):241–67.

60. Bühler A, Wang Q, Wolbers M, Model F, Beyersmann J. Comparison of Time-To-First-Event and Recurrent Event Methods in Multiple Sclerosis Trials. 2019;

61. Purroy F, Jiménez Caballero PE, Gorospe A, Torres MJ, Alvarez-Sabin J, Santamarina E, et al. Prediction of early stroke recurrence in transient ischemic attack patients from the PROMAPA study: a comparison of prognostic risk scores. Cerebrovasc Dis Basel Switz. 2012;33(2):182–9.

62. O'Connor P, Wolinsky JS, Confavreux C, Comi G, Kappos L, Olsson TP, et al. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. N Engl J Med. 2011 Oct 6;365(14):1293–303.

63. Freedman MS, Wolinsky JS, Comi G, Kappos L, Olsson TP, Miller AE, et al. The efficacy of teriflunomide in patients who received prior disease-modifying treatments: Subgroup analyses of the teriflunomide phase 3 TEMSO and TOWER studies. Mult Scler Houndmills Basingstoke Engl. 2018 Apr;24(4):535–9.

64. Tesfaw LM, Muluneh EK. Exploring and modeling recurrent birth events in Ethiopia: EMDHS 2019. BMC Pregnancy Childbirth. 2022 Aug 5;22(1):617.

65. Gabbett TJ, Ullah S, Finch CF. Identifying risk factors for contact injury in professional rugby league players--application of a frailty model for recurrent injury. J Sci Med Sport. 2012 Nov;15(6):496–504.

66. Confavreux C, O'Connor P, Comi G, Freedman MS, Miller AE, Olsson TP, et al. Oral teriflunomide for patients with relapsing multiple sclerosis (TOWER): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Neurol. 2014 Mar 1;13(3):247–56.

67. Steinman L, Fox E, Hartung HP, Alvarez E, Qian P, Wray S, et al. Ublituximab versus Teriflunomide in Relapsing Multiple Sclerosis. N Engl J Med. 2022 Aug 25;387(8):704–14.

68. Hauser Stephen L., Bar-Or Amit, Cohen Jeffrey A., Comi Giancarlo, Correale Jorge, Coyle Patricia K., et al. Ofatumumab versus Teriflunomide in Multiple Sclerosis. N Engl J Med. 2020 Aug 6;383(6):546–57.

69. Olesen AV, Parner ET. Correcting for selection using frailty models. Stat Med. 2006 May 30;25(10):1672–84.