14 April 2020
EMA/365402/2019

Overview of comments received on 'Treatment effect measures when using recurrent event endpoints qualification opinion' (EMA/CHMP/SAWP/291384/2019)

Interested parties (organisations or individuals) that commented on the draft document as released for consultation.

Stakeholder no.	Name of organisation or individual
1	AstraZeneca
2	EFPIA and EFSPI
3	EORTC
4	European Association of Hospital Pharmacists (EAHP)
5	International Society for Clinical Biostatistics, ISCB
6	Norman Stockbridge, H.M. James Ming Hung, Sue Jane Wang; Division of Cardiovascular and Renal Products and Division of Biometrics I, US FDA/CDER
7	Professor Jennifer Rogers, PHASTAR
8	Professor Stephen Senn, Consultant Statistician, Edinburgh
9	Regeneron Pharmaceuticals

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

1. General comments – overview

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
1	1. It would be helpful if it could be made clear in the Qualification opinion that clinical considerations also need to be taken into account when specifying the primary endpoint of interest in studies in this area. When studying subjects with risk factors for, but without established disease, e.g. heart failure (HF), the primary purpose of treatment is the prevention or delay of onset of HF. Therefore, we are primarily interested in delaying the first hospitalization for HF (hHF) as far as possible, since the occurrence of the first hHF is a good proxy for onset of HF. In this case, time to first hHF probably is a more appropriate primary endpoint than total number, or rate, of hHF. On the other hand, if we study subjects who already have established HF, then the therapeutic goal should be reducing the total number of hHF. In this case the rate of hHFs, including recurrent event (RE) analyses thus needs to be weighed in and serve as a more appropriate primary endpoint for capturing the effect on the full burden of disease. In the case above, when studying treatment effects in subjects who already have an established condition like HF, there is a natural need for including cardiovascular death (CVD) together with the less severe, and possibly recurrent event (hHF) as part of a composite endpoint. As stated above, recurrent event analyses better capture the full burden of disease but, combining morbidity with mortality increases the complexity. If, for example, an hHF event substantially increases the risk of CVD, the main treatment objective should be prevention of this hHF. In a	1. - Based on the comments, a statement clarifying different aspects of first event analyses and recurrent event analyses (cumulative effect over time) and the potential of both analyses to complement each other has been included in the document. Reference to the publication by Rauch is included. - It is agreed that in case of a disease like COPD or diabetes mellitus (hypoglycaemia) where an increase in recurrent less severe events has not a major impact on terminal events, recurrent event analyses are appropriate to measure a treatment effect. Astra Zeneca states that for disease prevention, which may include as well early stages of a disease, first event analyses are more appropriate, whereas in advanced stages recurrent event analyses are more appropriate to cover total disease burden. Although such a differentiation has its merits, a categorical difference between

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	population with HF and reduced ejection fraction there is such an increased risk for CVD after hHF and thus a time to first event analysis of the composite would seem most appropriate as primary analysis. In contrast, if an increase of the recurrent less severe events does not impact the risk of death, as is the case with exacerbations in COPD studies, a recurrent event analysis, including the total number of exacerbations and death would be a more appropriate choice of primary endpoint. Recurrent event analyses and time to first event analyses in general measure different treatment effects, and thus cannot be directly compared. These analyses do complement each other and carry different weight and importance depending on the clinical situation studied. It is our opinion that this should be made clear in the Qualification opinion. TTE analyses measure the direct (see paper by Rauch (reference below)) effect of the treatment whereas RE analyses measure the cumulative (direct and indirect) effect over time. It would be helpful if this distinction could be made clear in the Qualification opinion. 2. There are model selection issues associated with RE analysis. One general problem is that all recurrent events are considered instantaneous (to be able to model them using point processes, like, for example, in Andersen-Gill model, negative binomial or LWYY), hence, for example in CHF studies the duration of hospitalization is not taken into account. Since subjects being in the hospital are not at risk of having another HFH, then not considering the duration of hospitalizations can introduce biases in the estimation. This is highlighted in a recent paper by Lee and Cook (reference below). 3. For the patient -weighted approach it is noted that the estimation method is considered before first defining in advance the estimand of interest. There are no known models that use patient-weighted approach in the analysis.	these entities may not be generally valid for patients at increased cardiovascular risk. The differentiation between primary and secondary prevention is only one of several factors that determine the risk for cardiovascular events. Disease burden as determined by recurrent events can as well be relevant in a high-risk primary prevention population. Furthermore, it has been discussed controversially whether recurrent event analyses may be a tool in particular in early stages of diseases to increase the number of events. This may have an impact on the sample size required for pivotal studies. For these reasons it is preferred not to provide a general statement on the preferred application of first event vs. recurrent event analyses in prevention vs. later stages of the disease but instead refer to the "Guideline on clinical investigation of medicinal products for the treatment of chronic heart failure CPMP/EWP/235/95, Rev.2". The guideline states "Recurrent events are also important as they represent a large burden to patients".

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	References Rauch, Geraldine, et al. "Time-to-first-event versus recurrent-event analysis: points to consider for selecting a meaningful analysis strategy in clinical trials with composite endpoints." Clinical Research in Cardiology 107.5 (2018): 437-443. Lee, Jooyoung and Cook, Richard. "On estimands arising from misspecified semiparametric rate-based analysis of recurrent episodic conditions" Statistics in Medicine 2019: 1-22.	To which degree recurrent event analyses are more appropriate to capture a treatment effect is subject of the discussion of this document. 2. The duration of hospitalisation is discussed in further comments below. 3. This is acknowledged in the opinion.
2	The stated purpose of this application is on use of recurrent event endpoints in clinical trials. The opinion includes a long discussion of statistical approaches to analysis of such endpoints. It is surprising that these recurrent endpoint endpoints and analysis approaches were considered "novel" as they have been used in multiple submissions. Because of EMA considering this issue as a novel method for qualification, only confidential discussion with a single company has taken place. In the future we hope the development of statistical methods can be raised as a non-confidential cross-industry topic allowing for a broader range of opinions to be debated and considered e.g. at scientific meetings. The title of this opinion refers to recurrent event endpoints in general, but the actual opinion primarily discusses examples where reduction in mortality is a	This is acknowledged. However as mentioned in the text in cases where a terminal event limits the observation of the recurrent event endpoint novel approaches for recurrent events analysis are required. The Applicant was a consortium. Any reference to a single company was erroneous in the draft for publication. In addition, the draft and final qualification opinions are made public so that other researchers can comment and use the information or even built on it. This is acknowledged however also examples with no terminal events are discussed and as such the title will remain broader.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	primary goal of treatment and not those trials where mortality is very low. This should be reflected in the title of the opinion.	
	Much of the discussion on methods is based on a simulation study provided by the applicant. Conclusions based on simulations are entirely dependent on the assumptions used in the simulation model. More weight needs to be given to theoretical properties rather than basing conclusions on a limited set of simulations.	Agreed and mentioned in the final opinion.
	Abbreviations should be consistently throughout the document. In the first half of the document the abbreviation HFH is used for heart failure hospitalization. In the second half the abbreviation HHF is used without further explanation.	Abbreviations have been made consistent. In order to make clear that worsening of heart failure events may include both hospitalisation for heart failure and well defined outpatient emergency visits the term Heart Failure Event (HFE) is used throughout the document. The term "Hospitalisation for heart failure/heart failure hospitalization" (HFH) is used if this definition was used in a clinical trial.
	The issue on having a sufficient sample size to assess mortality does not seem to be a lot different for time-to-first event and recurrent event endpoints, although the latter might lead to lower sample sizes. It could be addressed by powering a trial to exclude a certain detrimental effect on mortality.	From a theoretical point of view this is true. In practice it turned out that by using a composite of death/CV death and HFEs (first events) the data needed for a mortality assessment have been provided in the past. The issue is discussed in detail in this document since this may not be the case if sample sizes become smaller with an

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
		estimand based on recurrent HFEs. When powering a trial to exclude a certain overall detrimental effect on mortality, issues like differential results in subgroups that have been relevant in the past in this therapeutic area have also to be taken into account. Just predefining an acceptable upper limit of a confidence interval may not be appropriate. The document does not aim at replacing the "Guideline on clinical investigation of medicinal products for the treatment of chronic heart failure" (CPMP/EWP/235/95, Rev.2) and no final conclusion on how a detrimental effect on mortality can be excluded in this specific therapeutic area is included here beyond the specific aspects that are discussed.
	There is some discussion of clustering of events of time, but this is not revisited when discussing methods. Some content around the impact of this on methods that assume marginal (or conditional on a frailty) rates rather than conditional risks is warranted. Approaches such as collapsing recurrent events into episodes of care or multistate models with differing risk conditional on an event are possibilities.	The qualification opinion encourages further methodological work, extension of the simulations etc. to include additional statistical approaches. It is noted, however, that also these approaches do have acknowledged limitations e.g. with respect to the number of parameters to be estimated.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	There is no discussion of non-proportionality and the interpretation and behaviour of recurrent event methods in this case. This problem is not unique to recurrent events versus a time-to-first event approach, but the impact and interpretation differs. For example, it is common for a delay from study to start before a treatment is effective. A recurrent event method will have less attenuation in this case than a time-to-first method possibly capturing the long-term treatment effect better.	The document does not only aim at analysing the statistical methods but heads for an integrated approach that includes the relevant clinical aspects to be considered. The possible reduction of study size by using recurrent events analyses was a key issue that has been discussed in the context of chronic heart failure in the past. In order to address the disease specific considerations it is important to bring up the implications of the approach for studies in chronic heart failure. The document deliberately focusses on one specific disease entity with the intention to specifically elaborate not only from a statistical but also from a clinical perspective what has to be considered when applying recurrent events. Clinical implications in other diseases may be different but have to be analysed with a similar level of detail. Therefore, addressing the implications on data on mortality has been an important integral part of the discussion in chronic heart failure. Now mentioned: if a valid and informative estimate for the recurrent event can be provided, this is seen as useful in a two-step

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
		approach including first an assessment of mortality. The qualification opinion is not restricted to methodological considerations but addresses recurrent event analyses also from a clinical and decision making perspective.
	There are clear limitations with the exposure-weighted approach as described in the document based on EMA's evaluation. The exposure-weighted estimand changes with the effect on the terminal event, but also changes with the duration of follow-up (and meaning interpretation would also need to consider changes in the study design). In addition, there is a loss of type I error for the individual assessment of the treatment effect on the recurrent event in situations, where the global null hypothesis is not true and the treatment effect regarding mortality is not neutral. EMA concluded that use of an approach for the recurrent event analysis where patients are given equal weight in the analysis regardless of the duration of follow-up may have potential to achieve this objective by recognizing the limitations. In addition, EMA concluded that there are also currently no established methods in the literature which target this estimand.	Comment acknowledged. No action needed.
	The applicant did not put forward a joint frailty model as an approach for estimating the patient-weighted rate in the presence of non-independent censoring. It is not clear why this was not evaluated by the applicant. The resulting rate estimator is conditional on the frailty but may match the reality of the problem and conditional and marginal rates often coincide. Moreover, a method exists for estimating the marginal hazard rates from such models (see Toenges,	The final qualification opinion mentions that alternative methods are available but because they were not extensively evaluated in this submission, they are not elaborated further.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Jahn-Eimermacher, Marginal hazard ratio estimates in joint frailty models for heart failure trials in Biometrical Journal, June 2019). In addition, EMA did not comment such existing methodology.While noting that the applicant did not provide alternative approaches, it is worth noting that the literature does contain possibilities and refer to joint frailty models not being reviewed in the correct qualification. EMA may want to comment on this. Comment: SAWP does not distinguish usefulness of recurrent event as endpoint in itself, from the side effect that use of recurrent event could displace mortality and result in less complete picture of mortality. Proposed changes: omit references to possible side effect of recurrent-event endpoint leading to smaller trials which in turn could lead to studies unable to detect differences in mortality; except to state this consideration and to note that this consideration is outside the scope of a Qualified opinion on recurrent event as endpoint.	The value of recurrent events endpoint are acknowledged but cannot be disconnected from the reality of clinical trials where terminal events are important, see further discussion in the opinion.
3	- Should this guideline be applicable for all diseases (chronic, non-chronic including cancers) or should this stay with some specific/predefined (chronic instead of acute, curable instead of terminal) diseases only? - In oncology, in adjuvant setting, diseases recurrence is an important event for disease free survival (along with death). This endpoint is broadly acceptable despite the lack of established surrogacies with Overall Survival (0S) in some cases. Recurrence by itself is rarely used to define an efficacy endpoint. Should this rule is applicable for other diseases? - Should surrogacy be required/established to OS for the recurrence event endpoints?	The document deliberately focusses on one specific therapeutic area, i.e. heart failure. In part since this was the example that was discussed in detail during the procedure. In addition, there has been a broader discussion about the applicability in heart failure studies. Focussing on one disease entity only allows to analyse in detail disease specific aspects, whereas extending the discussion to other disease entities would come at the price of a more superficial approach. Disease specific aspects concerning e.g. relevance of recurrences,

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
		treatment decisions and prognostic value are different in oncology. The same holds true for surrogates. This does not rule out the applicability to other diseases but implies that disease specific aspects have to be analysed in detail. A general statement concerning other diseases is added in the document.
	Recurrent events may not be independent of each other. This is shortly hinted at in the text when discussing the clinical background of recurrent HFH events: "Once hospitalized for heart failure, the rate of recurrent HFH is much higher." In clinical practice, for many types of recurrent events, the hazard of any given new event can be dependent on the history of the patient in a complex model. The risk can increase with each new event, decrease over time as long as no new event occurs or there may even be a 'quarantine' period after each event during which events are either impossible or very rare. The current text does not address these issues. Neither does it stipulate what assumptions need to be verified in order for the proposed methodology to be applicable and unbiased. It would be helpful if the opinion would also touch upon the issue of dependency between events.	Regarding the clinical aspects available data have been sited to highlight the issue. The aspect has been made more explicit in the assessment of the proposed methodology.
	Comment: The estimands developed in this proposal (Section 2.3) are based on rate without taking into account time. Why not considering the time to event approach (such as recurrence free or disease free) which will likely to offer a better	This aspect has been made in many comments and is agreed with respect to the question, whether an increase in the number

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	power for comparison and also offer more information as time component is added in addition to the recurrence event? Proposed change (if any): Use time to event estimands such as hazard ratio	of events always must be more efficient. Proposed estimates, however, take the time under observation into consideration
4	To the question whether "The measure of the treatment effect can be defined based on recurrent event endpoints" the CHMP produces an answer. The committee points out that recurrent event endpoints are well established in studies where the rate of terminal events is very low and reduction in mortality is not a primary goal treatment such as in relapses in multiple sclerosis and asthma. In cardiovascular studies is a different story. Assessment of both all-cause of mortality and cardiovascular mortality is mandatory. Based on a numerical scenario the CHMP points out that if the exposure-weighted rate is taken into account where the duration of the follow up of the patient is the same for all the patients, the patient-weighted rate method produces the same results of the exposure-weighted one. If the follow-up period is different for each patient, the results of the two methods will be different. The conclusion of the CHMP is that in studies where there are "no terminal effects" the methodology proposed (recurrent event endpoints) provides treatment effect measures that are more efficient than those based on first event only. Further, the CHMP recommends that decision-making analysis would consider estimates that summarize the expected effect of the treatment on the annual event rate while the patient is alive and an effect on the terminal event. The reasons of the CHMP are acceptable and well documented.	Comments acknowledged. No action needed.
5	I acknowledge the thorough investigations and discussions of the Academic Consortium regarding the value and limitations of different treatment effect measures and the corresponding statistical methods for the analysis of recurrent events. Furthermore, I welcome the effort of the Scientific Advice Working Party to	Comments acknowledged. No action needed.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	provide scientific advice on the definition of clinically interpretable treatment effect measures and suitable statistical analysis methods for recurrent event data. From an HTA point of view, the 2 most important limitations of the draft qualification opinion are given by the following. Firstly, methods for competing risks and methods for a joint analysis of the terminal event and the recurrent events are not considered in the scenarios with terminal event (Ghosh & Lin, Biometrics 2000; 56: 554-562 / Cook. & Lawless, Stat. Methods Med. Res. 2002; 11: 141-166 / Rogers et al., Stat. Med. 2016; 35: 2195-2205). Lately, efforts are being made in Germany to improve the statistical methodology applied to the analysis of adverse events in clinical trials. One of the points, which should be improved, is the consideration of competing risks and the application of suitable competing risks methods for time-to-first-event endpoints (Unkel et al., Pharm. Stat. 2019; 18: 166-183.). It is inconsistent to apply adequate competing risk methodology for the analysis of time-to-first-event endpoints, but to neglect this important issue in the analysis of the recurrent-event endpoints. Secondly, it is not mentioned that a thorough analysis of recurrent events should not be provided for one selected endpoint only. It should be added that an analysis of recurrent events should be performed for all relevant endpoints of this type to enable a meaningful and fair decision making	This point is well taken and may be the basis for a future EMA-driven initiative in the context of the development of a new guideline or the revision of an indication specific guideline. Development and application of competing risk methodology and multi-stage models require the development of substantial insight regarding the ability to use parsimonious models, which are fit for purpose and able to identify (and describe) the effect of an experimental treatment. A mere increase in the complexity of modelling is of no value in itself: models need to be well understood and usable for practical decision making. Additional efforts need to be made in how far such more complex multi-stage multi-parameter models provide information that is e.g. generalizable to a sufficient degree. Nothing in the current text specifically implicates that the positions as outlined, only refer to endpoints that are supposed to be primary.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	One general formal comment: Some tables and figures copied from the papers of the Academic Consortium have very bad quality. In part, they are unreadable.	Agreed. Better tables/figures will be provided.
6	Frequency of HFH plus CVD should be a meaningful indicator for disease burden, though it might not be good enough. Total duration in hospital seems also an important indicator for disease burden. It may be possible to create an order of importance considering both frequency and total duration in hospital together with CVD, and then apply Win-ratio or Finkelstein-Schoenfeld statistic for treatment comparison.	One of the issues with HFH is that it does not only represent severity of disease but also to some degree incorporates a decision based component and depends on availability and standards of the local health care system. This is by far more an issue for duration of hospital stay. The proposal to incorporate this component in the primary analysis is therefore not supported. Instead, broadening the HFH component to some by including very strictly defined urgent heart failure visits that do not lead to hospitalisation but reflect similar degree of deterioration is accepted in order to account for regional differences in health care system and to avoid missing events. This is in line with the current version of the CHMP heart failure guideline. In order to clarify this issue, the term HFH (hospitalisation for heart failure) has been changed to worsening of heart failure events (HFEs) throughout the text.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	If the treatment and the control yield equal time to death and time to censoring, then the exposure-weighted event rate seems more sensible than the equally-weighted event rate, for the reasons given by Consortium. If time to death or time to censoring differs between the treatment and the control, both the exposure-weighted event rate and the equally-weighted event rate are problematic to treatment comparison for HFH alone, but it may be okay for HFH+CVD.	No action needed.
	Recurrent event analyses present a better picture of outcome burden than do first events, which is good. However, no analysis that considers two one-day hospitalizations as worse than one seven-day hospitalization is entirely sensible. You are very likely to get more events if counting recurrent ones, but whether you get more power depends on your theory of the drug action. If you think the drug affects events regardless of when in the window of observation, then power should be better. But if you think the drug affects mostly early events and the later events are apt to be "different" somehow, then power may be decreased.	Regarding duration of hospital stay see also above. The comment concerns the issue of weighting events. In first event analyses weighting of HFH vs. death is ignored in the primary analysis, HFH is considered a categorical event indicating deterioration of disease. Weighting becomes more of an issue when recurrent HFH events are counted. The weight of an individual patient with several events increases as does the weight of HFH vs. death. The proposal to estimate severity of HFH events by length of hospital stay further complicates the issue. It is concluded that exploratory analyses calculating overall duration in hospital may be of interest to get additional information on disease burden but it should not be

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
		considered to be part of the primary analysis.
	The primary goal of an endpoint is to establish whether there is a treatment effect. Yes, it is nice if you can immediately translate the difference as one treatment clearly better than the other (and I am perfectly happy with prespecified weighting), but we will always decompose any composite endpoint to look at the effects on the components before concluding what was affected and whether there was net benefit.	Comment acknowledged. No action needed.
	Comparison of methods performed in the simulation studies may not be appropriate. Note that WLW targets marginal treatment effect, a challenge of its applicability in clinical trial for establishing treatment effect in confirmatory trials. Similarly, PWP may not target for an overall treatment effect parameter.	It was clarified in the final opinion that comparisons are based on simulated studies
7	Mortality There is a focus on the loss of information on mortality if studies are powered for recurrent events only. It should be noted, however that clinical trials of cardiovascular disease are currently predominantly designed based on a primary composite of first hospitalisation and CV death. The event rate for this composite outcome is commonly dominated by hospitalisation data and so trials are currently not powered on mortality. That said, I do appreciate that trials with recurrent hospitalisations as the primary outcome will typically be even smaller. The analysis of mortality can be accounted for in the analysis of recurrent events, either through a composite of recurrent events and mortality, or through the joint frailty model. Any analysis of a composite endpoint must also analyse, separately, the component parts to assess whether any treatment effect in the composite endpoint is consistent throughout. Consider an example where treatment has a large positive effect on time to first hospitalisation, but a negative effect on mortality. Analysis of the composite endpoint, dominated by hospitalisation events,	Whereas it is agreed that currently most studies are powered on a composite endpoint and not on mortality on its own, it is noted that based on this strategy in the past the data needed for a thorough assessment of mortality in the overall population as well as in relevant subgroups have been generated within such studies to either demonstrate superiority or at least exclude a negative effect. As outlined in the document there is a concern that this may not be the case if studies using recurrent events become smaller. For this reason, the document refers to what kind of mortality data is needed in chronic heart failure. It is

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	could mask the negative effect on mortality. So even in the current landscape, it remains vitally important that mortality is analysed as a stand-alone outcome as its inclusion in composite endpoints can often be found to be meaningless. Analysis of recurrent events as a primary outcome in clinical trials does not mean that a thorough investigation of mortality needs to be excluded. Recurrent HFH events The two statements: "Recurrent hospitalisations represent a considerable disease burden in patients with heart failure" and "The clinical meaningfulness of recurrent pre-fatal HFH events beyond a statistical booster of mortality remains to be clarified" appear to be at odds with each other. If HFHs remain a burden to patients, surely this is justification alone for their analysis, especially when, as quote, "mortality rates in CHF have decreased over the decades". Furthermore, even if they are only a statistical booster for mortality, surely that is reason enough to analyse them thoroughly. And whilst there are challenges associated with the analysis of HFHs, I do not believe that this should preclude their analysis.	concluded that at the end the size of the study may not be driven by a composite primary efficacy outcome using recurrent heart failure events but by the requirements on information on mortality needed. Agree, hospitalisations represent a disease burden for patients irrespectively of the time of occurrence. The statement has been amended. Clustering of such events in a single person just before a fatal outcome may unduly overweigh outcome in this patient. It is not clear that visiting a doctor shortly before deceasing indicates a worse clinical course as compared to dying without preceding hospitalisation events.
8	**Declaration of interest**: I know many of the consortium applicants (and some of the EMA CHMP experts) and have consulted for some companies the consortium members work for. I maintain a full declaration of interest here: http://www.senns.demon.co.uk/Declaration_Interest.htm ***1.1. Executive summary of my comments*** Many different factors impact on choice of endpoints for clinical trials. It is difficult to choose one analysis that will satisfy all needs. The argument that mortality *cannot* be combined with heart failure hospitalisations (HFHs) in a recurrent	Agreed. This is stated more clearly in the opinion

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	analysis *because* one is *adding* unlike events but *can* be reasonably combined in a time to first analysis *despite timing* unlike events is unreasonable. Recurrent event analysis will rarely be adequate on its own but this can be maintained about any analysis. One possible use of recurrent event analysis might be as part of a procedure in which (say) mortality and hospitalisation could be subsequently examined separately if the recurrent analysis showed a 'significant' effect. The door ought to be left open to using this as a possible approach. ***1.2. Some points about choosing analyses*** **1.2.1. Purpose** There are least two major purposes to analysing clinical trials. **Causal,** establishing whether there was a difference between the treatments and **predictive**, trying to estimate what it would likely be in future populations(Senn 2004). For the causal purpose, if under a strict null hypothesis two treatments may be assumed identical in the sample of patients studied, it may be of interest to detect some general effect in order to falsify this hypothesis. If and when it has been agreed that this hypothesis no longer holds, further examination may be appropriate in order to attempt to examine the relevance of this finding. Composite measures can have a valid role for addressing the first purpose. Once two treatments are judged to be different it then becomes relevant to examine more closely how exactly they differ. Once one has moved to this question, composite endpoints may be less relevant.	This view cannot be supported as a general rule: Drug regulation needs to assure that there is a formal proof of efficacy in some relevant construct of primary endpoint and this has to include a discussion of relevance as well upfront the assessment of benefit/risk in a broader concept. The first step of assessment also needs to be specific about the properties of the treatment effect that has to be demonstrated. In addition, ICH-E9 mandates for a consistent concept.

	1.2.2. Estimating standard errors adequately Nevertheless, even if a given signal can be validly estimated for a given (possibly limited) causal estimate, it is still necessary to estimate its standard error appropriately. Possible approaches include a) the **summary measures** approach(Finney 1990, Senn *et al.* 2000), whereby the data are reduced to a summary per patient and the standard error is based on such summaries (the per patient approach discussed in the guideline is along these lines) b) a suitable **mixed model** (Brown and Prescott 2014)c) an otherwise **sufficiently parameterised marginal model** (for example, the Poisson distribution being a single parameter distribution is not robust for error estimation and the negative binomial, also referred to in the guideline, is preferable)(Keene *et al.* 2007) d) **re-scaling of standard-errors** (for example, this is often applied to Poisson regression as an alternative to negative binomial regression(Liu and Menjoge 2008)) e) The **general estimating equation** approach(Liang and Zeger 1986) (which can be regarded as employing elements of a) b) and d)). f) **Resampling approaches** which, if they can be employed so as to adequately reflect hierarchies and dependencies, can be applied to many of the above. **1.2.3. Combining information** However, the various approaches in 1.2.2. , each of which if applied correctly, can provide unbiased estimates of some quantity of interest and appropriate estimates of the standard error of such estimates, can provide very different amounts of information. If the observation time per subject is very similar, then summary measures and mixed models can provide very similar estimates and estimated standard errors(Senn, Stevens and Chaturvedi 2000). However, if the amount of information varies dramatically from patient to patient then this is not the case. Consider, for example, the situation where two estimates, each based on the same number of patients would, if only the sample sizes were large enough, each yield the same estimate. However, the first is based on much shorter follow up than the	No action needed. The theoretical argument about two endpoints estimating the same treatment effect based on different observation times "asymptotically" may even be less efficient than discarding one half of the patients points in addition to the fact that it is already an assumption that for both endpoints the treatment effect was the same. The point is well taken that it is to some counter-logic to accept time to a HFH

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	second. Combining the two estimates by equal weighting (proportional to the number of patients) would yield an estimate with variance given by $$\sigma^2_{combined} = \frac{1}{4}\left(\sigma^2_1 + \sigma^2_2\right),$$ • where σ^2_1, σ^2_2 are the variances of the first and second estimate respectively. If $\sigma^2_1 > 3\sigma^2_2$, then $\sigma^2_{combined} > \sigma^2_2$, so that simply discarding patients of the first type would yield a more precise estimator. This is, in fact, one of the arguments against using Type III sums of squares in analysing multi-centre trials (Senn 2007) (Ch 14). It would be absurd if a regulator who was prepared to accept trials with a short follow up or trials with a long follow up if only they were large enough, would object to efficiently combining information from patients that could have been in either. **1.2.4. Random differences** Where estimates are being combined that may be expected to estimate related but possibly different quantities, then the discussion of appropriate variances in section 1.2.3. might be judged naïve. Clearly the philosophy being used to discuss information combination is that of a **fixed effects** rather than a **random effects** meta-analysis and a random effects meta-analysis will tend to weight different sources of information more equally. However, even in the context of combining information from different trials it has to be admitted that a) the fixed effect approach is perfectly adequate and indeed suitable for the causal purpose in 1.2.1. (Rice *et al.* 2017, Senn 2000) and b) the fixed effect approach is in any case the only approach that can be used for one or even two trials. In other words, the	event as a component of a composite primary time to event endpoint, but being reluctant using the repeated information. There is agreement that the composite endpoint is complicated and deserves further consideration regarding its precise interpretation even if a valid estimate of the standard error is available in survival analysis. There is also agreement that modelling of the HFH-events is an opportunity, but as explained in the clinical background there seems to be no obvious model. A strategy including a separate assessment of death and the rate of HFH-events is a solution not discussed in the current qualification procedure. No action needed.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	argument of 1.2.3. is reasonable (at least) for one of the major purposes of a clinical trial. **1.2.5. Differences in follow up** Of course, differential follow-up raises various issues and forces one to consider, because it brings the matter to one's attention, the possibility of non-constancy of effect. However, first, this is less of a problem for testing a strict null hypothesis, which brings us back to the matter of purpose, referred to in 1.2.1. above and second, it is equally a problem for understanding any trial in which, the treatments being judged to be different, the possibility of non-constancy of a treatment difference (or ratio etc) is raised. Thus, it is not reasonable to maintain that exposure time is generally inappropriate for understanding risk and indeed in lines 80 & 81 the Draft Qualification Opinion refers to a rate in terms of events per 100 person years without further comment. Clearly, this can, on occasion, be a perfectly reasonable thing to do. Indeed, censored patients in any conventional survival analysis will contribute different amounts of information if their follow up differs. Thus, conventional and commonly used analyses *do* use exposure time. **1.2.6. Combining different types of event** Of course, in combining mortality and hospitalisation one is combining very different things. However, it is a mistake to imagine that this issue is finessed by considering time to first event. That strategy can be defended as a simple way of producing valid standard errors (rather in the spirit of the summary measures approach in 1.2.2.) but the counting process that yields it is still counting different events.	No action needed. This is agreed, in this aspect (as explained above) also the currently used endpoint of time to first event is problematic, the great advantage is methodological correctness in a robust setting.

1.3. Summing up

The basic situation is difficult. As the CHMP opinion rightly points out, death is a more serious outcome than hospitalisation. Again, as the CHMP opinion points out, recurrent events such as hospitalisation are extremely undesirable for patients. In an ideal situation, huge trials analysed using multi-state and competing risk strategies(Schmoor *et al.* 2013) would yield reliable time-dependent transition probabilities based on risk factors and one could produce individual predictions that a given patient (or representative) could use to inform choice of treatment(Hilden and Habbema 1990). This is difficult to implement and the question then arises as to whether recurrent event analysis might represent a useful part of an overall strategy for judging the effects of treatments. That being so, the summing up of the guideline is too negative and the possibility of using recurrent event analysis as *part* of a general strategy for analysis should be considered.

Agreed, and particularly pointing to the fact that the use of multi-state models will not necessarily lead to smaller clinical trials.

The opinion accepts possibility of using recurrent event analysis as part of a general strategy for analysis.

References

H. Brown and R. Prescott (2014) *Applied mixed models in medicine*: John Wiley & Sons.
D. J. Finney (1990) Repeated Measurements - What Is Measured and What Repeats. *Statistics in Medicine,* 639-644.
J. Hilden and J. D. F. Habbema (1990) The Marriage of Clinical-Trials and Clinical Decision Science. *Statistics in Medicine,* 1243-1257.
O. N. Keene, M. R. Jones, P. W. Lane and J. Anderson (2007) Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharm Stat,* 89-97.
K. Y. Liang and S. L. Zeger (1986) Longitudinal Data-Analysis Using Generalized Linear-Models. *Biometrika,* 13-22.
D. Liu and S. Menjoge (2008) Statistical analysis of chronic obstructive pulmonary disease (COPD) exacerbations. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology,* 1422-1423; author reply 1423.
K. Rice, J. Higgins and T. Lumley (2017) A re-evaluation of fixed effect (s) meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 205-227.
C. Schmoor, M. Schumacher, J. Finke and J. Beyersmann (2013) Competing risks and multistate models. *Clin Cancer Res,* 12-21.
S. J. Senn (2004) Added Values: Controversies concerning randomization and additivity in clinical trials. *Statistics in Medicine,* 3729-3753.
S. J. Senn (2000) The many modes of meta. *Drug Information Journal,* 535-549.
S. J. Senn (2007) *Statistical Issues in Drug Development*, Hoboken: Wiley.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	S. J. Senn, L. Stevens and N. Chaturvedi (2000) Repeated measures in clinical trials: simple strategies for analysis using summary measures. *Statistics in Medicine,* 861-877.	
9	Regeneron welcomes the initiative by the Agency in releasing this 'Draft qualification opinion of clinically interpretable treatment effect measures based on recurrent event endpoints' and appreciates the opportunity to provide comments. The draft qualification opinion is a good first step in the formal evaluation of the application of statistical methods based on recurrent events for the purpose of regulatory submissions. The recurrent events, sometimes also referred to as multivariate survival endpoints, have been used in regulatory submissions and played critical roles in regulatory decisions, even though not pre-specified as the primary endpoints. We commend the Agency for publishing a draft qualification opinion on this relevant topic. Regeneron believes that the introductory section of this draft qualification opinion presents an opportunity for the Agency to highlight that, in addition to reductions of mortality, functional improvements should also be a main therapeutic goal in heart failure (HF). We understand that the discussion on the importance of mortality as a HF endpoint is needed to support the Agency's conclusions on this qualification request. However, the importance of functional improvements in this indication is increasingly recognised by both physicians and other regulators , and Regeneron believes it would be appropriate to acknowledge this in the "Clinical Background" section of this qualification opinion. At present, EMA's conclusions on the appropriateness and the utility of the recurrent event methodology appear to be mostly focussed on its application in HF drug development. Regeneron believes this methodology could also prove useful in other indications, both in the cardiovascular space and beyond, and we would	Comments acknowledged. No action needed. It is not the intention of this document to replace the "Guideline on clinical investigation of medicinal products for the treatment of chronic heart failure" (CPMP/EWP/235/95, Rev.2). In this guideline some reference to functional capacity is included. Focussing on one disease entity only allows to analyse in detail disease specific aspects, whereas extending the discussion to other disease entities would come at the price of a

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	welcome the Agency to more openly acknowledge this. Stakeholders would benefit from the Agency's recognition that other types of recurrent events or multivariate survival endpoints have wide application in understanding the clinical effect of treatments. Some examples include, but are not limited to: events of stroke, myocardial infarction, unstable angina requiring hospitalisation, or also the occurrence of various types of opportunistic infections in AIDS clinical studies; such multivariate survival endpoints are often considered as recurrent events with proper clinical definition and adjudication. Lines 180-183 briefly mention the utility of the recurrent event approach beyond HF drug development, but a further discussion in the "Conclusion – qualification opinion statement" section would be beneficial. Regeneron recognises that this qualification opinion has a specific scope and is based on the assessment of data submitted to the EMA. However, this opinion also allows the Agency to expound on its stance regarding the use of the recurrent event methodology in other chronic indications where new statistical approaches could support an accurate evaluation of treatment effects, and potentially help facilitate drug development. In addition to being more inclusive and expanding the scope of this statistical approach to other clinical applications, we would also encourage the Agency to further discuss the interpretations of various statistical methods and their properties in this qualification opinion. It is our position that this opinion would benefit from the inclusion of other potential statistical methods and commensurate interpretation, and that these could clarify stakeholder understanding of the Agency's expectations.	more superficial approach. Methodological concepts elaborated in this opinion could be adjusted to other diseases. Comments acknowledged. The opinion includes already a discussion on interpretation and properties of the proposed statistical methods.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Apart from the methods evaluated in this draft qualification paper and the Agency's response, there are other relatively new and advanced statistical methods available in literature. Alternative methods and approaches should be considered, and their utility and limitations discussed within this document. For example: • joint modelling the recurrent event and terminating events has been seen in recent publications; • the frailty model, which introduces randomisation of effects for recurrent events, is also available. The discussion of additional challenges in the analyses of recurrent events would also prove useful and meaningful to Sponsors. Expanding this discussion to highlight utility in specific or representative indications/therapeutic areas would strengthen this draft opinion and could help guide investigators looking to use this methodology in the future. Examples of specific topics that might warrant discussion include: • Severity or sizes of the events may need to be considered: for example, the sizes of tumours or the severity of migraines may also carry information on treatment effects; • Duration of recurrent events: duration of migraine, duration of pain in addition to the frequency of occurrence may also need to be considered; • Correlation of terminating events with the recurrent events: it may be of interest to evaluate if higher rates of recurrent events may lead to increased risk of terminating events. Additional clarification around these topics may facilitate the implementation of this statistical methodology and benefit drug development.	The opinion does not expand to alternative methods because these were not extensively discussed in the submission. Regarding duration of recurrent events: One of the issues with HFH is that it does not only represent severity of disease but also to some degree incorporates a decision-based component and depends on availability and standards of the local health care system. This is by far more an issue for duration of hospital stay. The proposal to incorporate this component in the primary analysis is therefore not supported. It is agreed that mortality is not independent from recurrent HFH events. As has been described in the introduction clustering of

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
		HFH events before fatal events has been described.

2. Specific comments on text

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
29-34	2	It is not clear what constitutes a terminal event? If it includes relapses, exacerbations, etc. I suggest skipping "e.g. death" and the reference to mortality in the sentence that starts at line 29. In addition, the term "terminal event" seems not suited for events other than death.	Other examples for terminal events may be survived resuscitation, kidney transplant or heart transplant.
45-46	2	Comment: Line refers to time to first heart failure hospitalization with a reference to CPMP/EWP/235/95, Rev2, 20, July 2017. In that Guideline, it refers to an evolution towards more non-hospital setting care. That additional context is lost in the qualification report and contemporary trials are inclusive of such events. Proposed change (if any): Either change to time to first heart failure **event** or end the sentence at "related to worsening of heart failure" and refer to the Guideline.	Agreed. Strictly defined outpatient urgent visits for worsening of heart failure are accepted. The term is changed to "heart failure event (HFE)" throughout the document where applicable and an explanation is added: "e.g. time to first hospitalisation for worsening heart failure event (HFE) which may include hospitalisation for worsening of heart failure but also well-defined outpatient visits for worsening of heart failure." (Guideline on clinical investigation of medicinal products for the treatment of chronic heart failure (CPMP/EWP/235/95, Rev.2, 20, July 2017)).
56	2	Comment: Line refers to "The inclusion of recurrent events as co-primary endpoint may be considered" by quoting the reference to CPMP/EWP/235/95, Rev2, 20, July 2017 (the main therapeutic goals in the treatment of CHF are to reduce cardiovascular mortality and to prevent deterioration of the clinical status and hospitalizations; these	It is preferred to leave it as it is since it quotes the HF guideline. The guideline states:

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		goals should represent the primary aim of new agents developed for the treatment of CHF. Proposed change (if any): The inclusion of recurrent events as one of primary endpoints may be considered.	"The inclusion of recurrent events as co-primary endpoint may be considered, but this setting needs further justification, adjudication of the events and a clear methodological strategy".
65	2	Suggest using the abbreviation "CHF" consistently.	Agreed
67	2	Suggest using e.g. "treatment regimen" instead of "treatment algorithm".	Agreed
86	2	Comment: Referring to PARADIGM at 8442 subjects as "reasonably sized" could be viewed as providing sizing guidance, while this study is on the larger size of historic chronic heart failure (CHF) studies including the largest under some classifications of their studies population. Proposed change (if any): "The study is an example for a ~~**reasonably sized**~~ **large** study (8442 patients) able to provide the data needed for assessment of effects on mortality and hospitalization for patients as included in this study".	The size of the study is within the range of current cardiovascular outcome studies aiming at an application based on one pivotal trial that meets the expectations on a robust and compelling study. Since it is not necessary to comment on the size the term is changed to "… for a contemporary study".
96-99	2	Comment: The reduction in variability could also be used for a power gain instead of a decrease in sample size. For example, a trial could still be powered for the traditional time-to-first event endpoint, but recurrent events used as primary analysis. Proposed change (if any): Reduction in variability in estimates, mainly discussed from the background of an opportunity to reduce the overall sample-size of a trial may thus limit the opportunity of risk-benefit assessment in an indication that suffers from high unexplained variability that should be acknowledged. A potential alternative would be to use the reduction in variability for a power gain instead of a sample size reduction.	Reference to the Guideline on the investigation of subgroups in confirmatory clinical trials (EMA/CHMP/539146/2013) is added.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
102-103	2	Comment: Suggest qualifying "in subgroups" to limit to those with a hypothetical rationale for differential treatment-based safety or efficacy as it would be burdensome to design to avoid exploratory subgroup differential results. Proposed change (if any): "As a prerequisite the data have to provide sufficient reassurance that mortality is not increased to a relevant degree in the overall population and in ~~**subgroups**~~ **treatment and disease relevant subgroups**".	No change implemented.
115-117	2	Comment: The robustness of the subgroup results in ValHeft is controversial, for example it was not reproduced in CHARM-Added (White, 2003, Lancet) Proposed change (if any): In Val-HEFT, the neutral effect on mortality was the net result of a significantly increased mortality in patients receiving in addition ACE inhibitors and beta blockers, and a significantly decreased mortality in the other patients. However, it should be noted that this finding was not reproduced in CHARM-Added and remains controversial.	It is agreed that the ValHeft results have not been consistently reproduced. The study is not cited to indicate a demonstrated detrimental effect on triple therapy but to serve as an example for divergent effects in subgroups leading to an overall neutral effect on all-cause mortality. The text has been amended.
136	2	Suggest revising this sentence. As it stands it reads like: Hospitalisation causes recurrent HFH.	The sentence has been revised accordingly.
141-142	2	Suggest replacing "linearly" with "evenly".	The sentence has been revised.
143	2	Suggest clarification: What is a "Statistical Booster of Mortality"?	The paragraph has been revised to better explain the difficulties associated with the interpretation of pre-fatal clustering of hospitalisations. It cannot be generally assumed that it represents a worse clinical course if a patient has

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
			been in hospital or in an emergency some days or weeks before dying than if he dies at home without seeing a doctor.
149-158	2	Comment: While there are examples of trials with positive effect on HFH and neutral effect on mortality, like the Val-HeFT trial, and examples with neutral to small effect on hospitalizations and detrimental effect on mortality, like the referenced Xamoterol trial, there does not seem to be examples of trials with observed positive effect on hospitalizations and detrimental effect on mortality. These would be the real cause for regulatory concern, and it seems fair to mention that there are no trials with such observed effects. Proposed change (if any): HFH or signs and symptoms of heart failure did not exactly mirror the effect of a treatment on mortality in the above mentioned two studies with milrinone and xamoterol. Also the DIG study is an example of discrepant results for both parameters. However, at least it seems reassuring that currently no published trial shows an observed positive effect on hospitalizations and an observed detrimental effect on mortality.	It is agreed that no such robust examples are known when it comes to results of the whole group of patients. The assumption is less clear for subgroup analyses. Although not based on a predefined analysis, the post hoc results of a substudy in patient in the DIG study that are discussed in the document indicate the possibility that patients at higher glycoside levels had an adverse outcome for mortality despite of an positive outcome for the hospitalisation component in the overall group. Therefore, it is proposed not to change this section.
151-153	2	The Qualification opinion supports the point that "it cannot be assumed a priori for a new therapeutic agent that HFH [hospitalisation for heart failure] is predictive for mortality" with the following logic "HFH or signs and symptoms of heart failure did not exactly mirror the effect of a treatment on mortality in the above mentioned two studies with milrinone and xamoterol… Also the DIG study is an example of discrepant results for both parameters." Comments:	The text has been amended to address some of the concerns.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		1. Whether signs and symptoms mirrored mortality exactly is not relevant to an assessment of HFH as predictor of mortality effect. 2. That HRH "did not mirror" mortality effects is not precise or strong evidence against HRH as predictor of mortality effect. 3. The DIG results showed improved hospitalisation but no effect on mortality; the above text obscures this somewhat Proposed changes: Change "it cannot be assumed a priori for a new therapeutic agent that HFH [hospitalisation for heart failure] is predictive for mortality" to "HFH may not be predictive of mortality". Change "HFH or signs and symptoms of heart failure did not exactly mirror the effect of a treatment on mortality in the above mentioned two studies with milrinone and xamoterol… Also the DIG study is an example of discrepant results for both parameters" to "For example, the DIG study results estimated improved hospitalisation but no significant effect on mortality".	
164-166	2	Comment: The Qualified opinion includes as weaknesses of recurrent event endpoint "disease specific differences" and "factors like health care supply that may have an impact"; but such differences would tend to affect all endpoints. Proposed changes: Change "Among the challenges when clinically interpreting recurrent event HFH are disease specific differences, clustering of events and factors like health care supply that may have an impact on the event rate" to "A particular challenge to the clinical interpretation of the recurrent event HFH is the clustering of events; in addition, health	The text has been amended. The issue of health care supply and local preferences may be more relevant for the overall analysis if more recurrent events are counted in the primary analysis. The impact of disease specific differences (e.g. differences HFE rates in idiopathic vs. ischemic cardiomyopathy) on the overall result has to be analysed.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		care supply may have an impact on the event rate, although health care supply may affect other outcomes also."	
166-167	2	Comment: "Studies may become smaller when sample sizes are calculated based on recurrent HFH. This has a relevant impact on data available for mortality assessment". This is not an argument against recurrent event as endpoint, only against smaller trials. Proposed changes: omit the quoted text.	Agreed but the importance of this aspect should be acknowledged.
167-171	2	The composite of mortality and HFH seems not a suitable endpoint, not even according to existing ICH guidelines.	Agreed to delete since this is more a theoretical concern.
167-171	2	Comment: Risks of HF hospitalizations and mortality are generally observed to be highly correlated, see for example Table 2 of the Kristensen et al (2015) on risk groups of the I-Preserve study. Including mainly patients with a low risk of mortality would therefore also lead to the inclusion of patients with low hospitalization rates (first and recurrent), which is not desirable for a sponsor. Proposed change (if any): Delete sentence starting with "Moreover, …" as statement does not seem correct.	i) This refers to current knowledge to be taken for granted also for future situations. ii) non-inferiority studies should be considered, as well for the decision about the generalisability of certain statements.
170	2	Comment: Typo of "witch" Proposed change (if any): "…component of a primary endpoint may stipulate inclusion of patients at lower risk ~~**witch**~~ **which** may further decrease the robustness of information on mortality".	Sentence has been deleted.
189-229	2	This section implies there are only two methods for calculating rates within treatment groups.	It is noted that the qualification option is centred around the submitted proposal of the applicant. Attempts were made to be more specific about

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		For the "patient weighted" approach, the only approach considered is a simple arithmetic mean of rates. Rates will typically be skewed; a median or other summary statistic rather than an arithmetic mean will more closely represent the goal on line 225-226 of "What patient considering what annual rate they as an individual might expect while they are alive." For example, few would consider the arithmetic mean of everyone's salary to represent what an individual might expect in terms of their salary. Instead it is customary to use the median. The rate estimated from a negative binomial model often provides an appropriate compromise between the "exposure-weighted" and "patient-weighted" approaches. It has been stated multiple times at scientific meetings by the ICH E9 authors that the "summary" part of an estimand can be an estimate from a model. Therefore, it is misleading to focus only on two methods for calculating rates.	assumptions underlying the use of the different estimands. Please note, that we are intending to find clinically meaningful estimates and in no instances the qualification opinion is prescriptive. The qualification opinion clearly states that further methodological research is needed, going beyond simulations, but also addressing the need to communicate the outcome of the methodological calculations to the stakeholders of the research and its applications.
190-229 and ff	2	Comment: Distinguish patient-weighted from exposure-weighted counts of recurrent events. Patient weighted events do not standardise the number of events for a patient by the exposure of the patient. Despite the emphasis put on the patient-weighted measure in the Qualified opinion, it is doubtful that patient-weighted could be extensively used in practice because exposure may not be independent of treatment effect. Yet the Qualified opinion seems to favour this very difficult-to-interpret endpoint where number of events is not weighted by the time over which the events were observed, and where absorbing events such as death would have maximum impact on the targeted measure of recurrent events. Could a second opinion be obtained from a statistician by the drafters of the Opinion on the use of patient weighted events? The	Limitations of both approaches are mentioned now. The patient weighted approach is understood as an alternative to an established estimate that heavily depends on distributional assumptions that may not be fulfilled.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		lack of a denominator for the rate is constitutes good statistical grounds for omitting major consideration of the proposed patient weighted counts of occurrences. Proposed changes: Re-think discussion of patient weighted counts of recurrent events given their lack of a denominator and the importance of a denominator in this context. This section could include more discussion of events that are missing not at random (MNAR).	
250-257	2	Comment: Line 257 refers to "control 0.667 and ration 0.5" for patient weighted. The calculation is wrong – "control" should be 2.66667/5 = 0.533333 and ratio should have been 0.333/0.533333=0.62. Line 250: the HFF rate is ~0.62 on treatment compared to control on a per-patient basis (rather than half). Proposed change (if any): Please see the correct numbers stated above.	Treatment and control had 4 patient both. So, the denominator is 4.
258-264	2	Comment: Note that in the described situation using the recurrent composite endpoint (estimand 2) gives the more intuitive answer that treatment B is preferred. Counting CV death as event gives 4 events per year of follow-up while alive and treatment B 3.5 events per year of follow-up while alive. The analysis of mortality that would additionally be done would also clearly identify treatment B as the better option. Proposed change (if any): Add sentence "Consideration of estimand 2 gives the more intuitive answer that treatment B is the preferred treatment."	Agreed, please find also comments on the prerequisites of estimand 2 being Poisson distributed.
270 ff	2	Section 2.3.x It looks like that the term estimand and estimator are sometimes mixed. Please check against the terminology in ICH E9 Addendum.	Comment acknowledged.
272-274	2	Comment: The applicant did not put forward a joint frailty model as an approach for estimating the patient-weighted rate in the presence of non-independent censoring. It is not clear why this was not evaluated. The resulting rate estimator is conditional on	The qualification opinion is centered around the submission of the applicant.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		the frailty but may match the reality of the problem and conditional and marginal rates often coincide. Moreover, a method exists for estimating the marginal hazard rates from such models (see Toenges, Jahn-Eimermacher, Marginal hazard ratio estimates in joint frailty models for heart failure trials in Biometrical Journal, June 2019). Proposed change (if any): While noting that the applicant did not provide alternative approaches, it is worth noting that the literature does contain possibilities and refer to the joint frailty models not being reviewed in the correct qualification.	It has been added to the text that further methodology is available.
281-282	2	The statement that "LWYY is the Anderson-Gill method, which gives the same point estimate as negative binomial regression" is incorrect. They provide different estimates. See for example: Keene ON, Jones MR, Lane PW, Anderson J. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. Pharmaceutical Statistics. 2007 Apr;6(2):89-97.	Deleted.
293-295	2	For the simulation scenarios "it was assumed that after discontinuation from active treatment patients were followed up and event rates went back to the control rate." Although not proposed by the sponsor, it would be helpful for the Qualified opinion to reference Keene et al. (2014) Missing data sensitivity analysis for recurrent event data using controlled imputation, *Pharmaceutical Statistics,* which proposes an analysis that fits this assumption for recurrent events exactly – software is available at missingdata.org.uk.	The qualification opinion is centered around the submission of the applicant.
307-310	2	The conclusion that "there is a possibly a small loss of control with recurrent event methods" is not warranted and should be qualified. Statistical theory shows that if the assumptions of any method are met, then type I error will be controlled at the appropriate level. The values in table 1 depend on simulating data that depart from the assumptions of the models.	A more cautious wording has been used.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		For example, a similar table showing lack of control of type I error could be produced for a simple t-test by simulating data that did not meet the assumptions of a t-test.	
344-347	2	This line again refers to "an issue with type I error control for small sample sizes". This issue is related to the assumptions used in the applicant's simulations not to any wider problem with these methods.	Statement that evaluation is based on simulated data is added.
377	2	Comment: The described pattern, that a treatment with detrimental effect on mortality is preferred does not occur for estimand 2 in Table 8 and it also does not seem to occur tables in other scenarios presented in the application. While situations could potentially be constructed where such a pattern would be present for estimand 2, the magnitude of the effect would at least seem to be bounded. For low mortality rates the impact of a treatment effect on mortality is limited anyway. And for high mortality rates estimand 2 will favour a treatment with a positive effect on mortality, as shown on slide 23 of the Applicant's replies to the second list of issues. So for estimand 2 the described pattern could occur only with a limited magnitude for intermediate mortality rates. Proposed change (if any): This pattern does not occur so markedly with estimand 2 in the above tables and other presented scenarios.	This is true, particularly for small event rates as used in these simulations, the effect of dependencies will not be prominent.
377-379	2	Comment: A more thorough discussion of the behaviour of estimand 2 would have been expected here. It seems that many of the disadvantages mentioned by the CHMP for the exposure-weighted approach are overcome or at least alleviated by the use of estimand 2. That would include the pattern of dependency on effect on terminal event and the representation of patients with short follow-up time, who contribute an event with estimand 2. While mortality is certainly a worse event than a hospitalization, events of different severity are commonly combined into one composite endpoint, followed by an	In the concept as proposed (i.e. without a separate co-primary assessment of mortality) the estimate currently suffers from a clear interpretation. This contradicts the regulatory need to clearly describe which aspect of disease is changed by treatment. This is true. The situation is hampered by the fact that in the example of HFE

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		investigation of the effects on the components. CVD and HHF as first events would also have the same weight in a time-to-first composite event analysis, which is nevertheless applied and endorsed by the EMA guideline on the development of treatments for heart failure. The case of recurrent events does not seem to be fundamentally different, so giving the equal weight of CVD and HHF as sole reason for dismissing estimand 2 does not seem appropriate. One could even argue that at least all CVD cases are included in the estimand 2 analysis, which is not the case for a CVD after an HHF in a time-to-first event analysis. Proposed change (if any): Include further discussion of estimand 2.	traditionally mortality has been the primary endpoint. So the justification for the combined endpoint is not so clear as in other diseases.
445-446	2	"The (targeted) effect (when endpoint is exposure-weighted) also alters with other design properties such as the duration of follow-up". This statement needs to be justified – seems true only if rate varies over time in a systematic way; and in that case, both patient-weighted and design-weighted measures would be equally affected. Proposed change: Change "The effect also alters with other design properties such as the duration of follow-up." To "If event rates vary over time, the estimate of effect will depend upon the duration of follow-up of the trial, whether the events are patient-weighted or exposure-weighted."	We deleted this aspect.
Lines 80-82	5	Comment: The presented confidence intervals for the rates per 100 patient years for all-cause mortality, CV death and first HFH are probably based upon the assumption that the corresponding survival times are exponentially distributed. The validity of this assumption is questionable. A better option is given by the citation of the corresponding hazard ratios given in Murray (*New Engl. J. Med.* 2014; 371: 993-1004).	The section has been amended. The rate over a median follow up of 27 months as provided in the NEJM publication and not the rate per 100 PYs as provided in the published EPAR are now cited.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		Proposed change (if any): Replace the presented rates per 100 patient years for all-cause mortality, CV death and first HFH by the corresponding hazard ratios given in Murray (*New Engl. J. Med.* 2014; 371: 993-1004).	
Lines 258-259	5	Comment: The consideration of the frequency of HHF independently of mortality is inadequate and should not be done in practice. Proposed change (if any): Please add that the presented consideration is just a theoretical one used for explanation and that in practice the interpretation of HHF results independently of mortality should not be done.	We agree and this should be the basis for an assessment strategy to be outlined in future indication specific guidelines.
Lines 262-264	5	Comment: The presented "HTA-conclusion" obviously only takes costs into account and neglects the overall benefit-risk ratio for the patients. This is not in line with the general HTA view and should be rephrased. Proposed change (if any): Replace "HTA-conclusion" by the phrase "… the conclusion if only costs were considered …" or something like this.	Agreed.
Lines 267-268	5	Comment: It is correct that the independent interpretation of treatment effects for recurrent events and terminal events leads to obvious problems. Therefore, such an independent interpretation should not be done in practice.	The opinion provides suggestions for the clinical and statistical workout.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		Proposed change (if any): Please add that treatment effects for recurrent events should not be interpreted independently of terminal events in practice.	
Lines 311-313 Table 7a	5	Comment: In Table 7a Type-1 error rates are presented for 1-sided tests. The corresponding results for the usual 2-sided tests originally presented by the Academic Consortium are preferable. Proposed change (if any): Replace the results for the 1-sided tests by the corresponding results for the usual 2-sided tests.	One-sided type-1-errors directly reflect the error in decision-making for positive decisions.
Lines 353-354 and 453-460	5	Comment: It is correct that terminal events complicate the estimation of the reduction in recurrent events. I support the CHMP encouraging research for this data situation. However, methods for competing risks and methods for a joint analysis of the terminal event and the recurrent events are not considered. It is sensible to apply and extend methods for competing risks not only for time-to-first-event endpoints but also for recurrent-event endpoints considered here (Ghosh & Lin, *Biometrics* 2000; 56: 554-562 / Cook. & Lawless, *Stat. Methods Med. Res.* 2002; 11: 141-166). Another option is given by the joint analysis of the recurrent events and the terminal event to avoid a misleading interpretation of the results for recurrent events independently of the terminal event (Rogers et al., *Stat. Med.* 2016; 35: 2195-2205). Proposed change (if any):	The qualification opinion is centered around the submission of the applicant. The mentioned methods deserve further investigation.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		Please add that the application and extension of competing risk methodology for recurrent-event endpoints is useful and add a discussion of available methods for the the joint analysis of the recurrent events and the terminal event.	
Lines 434-460	5	Comment: I agree that the analysis of recurrent events is useful when the corresponding effect measures provide a better description of the patients' disease burden than the analysis of the first event only. However, for a meaningful decision making such analyses should not only be performed for one selected endpoint but for all relevant endpoints of this type. Proposed change (if any): Please add that an analysis of recurrent events should not only be performed for one selected endpoint but for all relevant recurrent-event endpoints to enable a meaningful and fair decision making.	See above.
Starting at 461	8	Comment: The summary is too dismissive of an approach that might be useful on occasion. An extra paragraph should be added Proposed change (if any): Add paragraph at end as follows: "Recurrent event analysis might on occasion be useful as an initial analysis, perhaps performing a 'gatekeeper' function. Given a positive result, further supporting analyses would usually be appropriate. Reassurance would be needed that standard errors and probability statements were being appropriately calculated. Questions of multiplicity would have to be properly addressed."	We would prefer the strategy to be centered around an assessment of mortality.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
Lines 180-183	9	Comment: The current wording of the draft opinion suggests that the recurrent event approach for the estimation of treatment effects in HF is mostly valuable in Phase 2 trials and extrapolation exercises. We believe that this approach could also prove useful beyond those scenarios, namely in Phase 3 trials in HF. Proposed change (if any): "Although not within the scope of this methodological qualification opinion, the application of recurrent HFH ~~in areas where robust data on mortality are less important (e.g.~~ could prove useful beyond phase 3 trials and also add value to phase 2 trials~~,~~ and extrapolation exercises (where robust data on mortality are less important), or in rare diseases, where information on mortality primarily depends on the number of patients available and not on the study design~~,~~; use of this methodology in these scenarios is endorsed by CHMP."	The points to be considered from a clinical perspective are summarized in the paragraphs above. The document focusses on analyses most relevant for decision making. The last paragraph provides proposals for additional possible applications of the method beyond what has been said for large scale phase 3 studies and for areas where it may be valuable to further explore the application.
Lines 282-284	9	Comment: We encourage the Agency to consider revisiting its statements regarding the WLW and PWP methods and the interpretation of the application of recurrent events. WLW models the marginal multivariate survival analyses and estimates the marginal hazard ratio for the first event, second events, etc. The estimation of WLW is different from NB and Anderson-Gill methods. The interpretation of WLW on recurring events is usually focussed on questions such as: whether a treatment is effective by delaying the first event, or whether it also has effect on the second or third event. The statement on the application of WLW when terminating event present could potentially be revisited. The discussion and interpretation of WLW when applied with multivariate survival events with terminating events are available in literature[1], which may provide additional approaches and considerations when analysing recurrent events with terminating events. The interpretation of the WLW method usually depends on how the	Two estimands have been proposed by the applicant and put into perspective with other methodology. It is true (and mentioned in the qualification opinion) that all these methods address different aspects of the "treatment effect". The need for further methodological research and attempts to better communicate the outcome of this research are welcome.

[1]Li QH, Lagakos SW. Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. Stat Med. 1997 Apr 30;16(8):925-40.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		terminating events are treated in the model, which could potentially be further discussed in this qualification opinion. The PWP method may have limitations in its application primarily due to the potential large risk set reduction for the analyses of subsequent events after the first event; limitations of this method are briefly discussed in this draft qualification opinion. However, the interpretation of the analysis of results using this method may potentially benefit from further evaluation and more nuanced discussion, given the limitations noted above. Proposed change (if any): N/A	
Lines 319-321 (Tables 5 and 6)	9	Comment: Regeneron suggests that the Agency reconsider the biases presented in Tables 5 and 6 for the WLW method. As previously noted in our comments above for lines 282-284, the WLW method does not estimate the parameters as the NB and LWTT do. When evaluating bias, it is necessary to calculate the true value that each method estimate. Therefore, we recommend that the Agency consider incorporating the statistical properties for all methods evaluated; including the assumptions, asymptotic properties, (e.g., consistency and efficiency), as well as properties such as bias, type I error and power when assumptions are incorrect. Proposed change (if any): N/A	No action.