EXTENDED REPORT

Development of EULAR recommendations for the reporting of clinical trial extension studies in rheumatology

Maya H Buch,1,2 Lucia Silva-Fernandez,3 Loreto Carmona,4 Daniel Aletaha,5 Robin Christensen,6 Bernard Combe,7 Paul Emery,1,2 Gianfranco Ferraccioli,8 Francis Guillermi,9 Tore K Kvien,10 Robert Landewe,11 Karel Pavelka,12 Kenneth Saag,13 Josef S Smolen,5,14 Deborah Symmons,15,16 Désirée van der Heijde,17 Joep Welling,18 George Wells,19 Rene Westhovens,20,21 Angela Zink,22 Maarten Boers23

ABSTRACT

Objectives Our initiative aimed to produce recommendations on post-randomised controlled trial (RCT) trial extension studies (TES) reporting using European League Against Rheumatism (EULAR) standard operating procedures in order to achieve more meaningful output and standardisation of reports.

Methods We formed a task force of 22 participants comprising RCT experts, clinical epidemiologists and patient representatives. A two-stage Delphi survey was conducted to discuss the domains of evaluation of a TES and definitions. A ‘0–10’ agreement scale assessed each domain and definition. The resulting set of recommendations was further refined and a final vote taken for task force acceptance.

Results Seven key domains and individual components were evaluated and led to agreed recommendations including definition of a TES (100% agreement), minimal data necessary (100% agreement), method of data analysis (agreement mean (SD) scores ranging between 7.9 (0.84) and 9.0 (2.16)) and reporting of results as well as ethical issues. Key recommendations included reporting of absolute numbers at each stage from the RCT to TES with reasons given for drop-out at each stage, and inclusion of a flowchart detailing change in numbers at each stage and focus (mean (SD) agreement 9.9 (0.36)). A final vote accepted the set of recommendations.

Conclusions This EULAR task force provides recommendations for implementation in future TES to ensure a standardised approach to reporting. Use of this document should provide the rheumatology community with a more accurate and meaningful output from future TES, enabling better understanding and more confident application in clinical practice towards improving patient outcomes.

INTRODUCTION

A randomised controlled trial (RCT) is the most objective means of evaluating an intervention and underpins regulatory decision-making and, if appropriate, the introduction of therapies into clinical practice. Many benefits of RCTs have been seen in the specialty of rheumatology, and particularly in the management of rheumatoid arthritis (RA).1–10 While the aim of RCTs is to demonstrate the efficacy and safety of an experimental agent, their observation period typically spans a relatively short time frame. However, the use of therapies in chronic diseases necessitates more long-term evaluation. The introduction of new disease modifying anti-rheumatic drug (DMARD) therapies for the treatment of RA has been associated with a significant number of post-RCT extension studies.13–17 henceforth termed ‘trial extension studies’ (TES), to report the longer-term outcomes of an experimental agent.

Role of TES

TES can evaluate in particular, the effects of cumulative exposure to a drug, capturing events through systematic reporting, monitoring of source data, and consistent coding, thus enabling further assessment of the long-term safety profile observed during the RCT.18

An additional benefit of TES that is cited is continued access to an effective but otherwise unlicensed treatment by RCT participants. However, since a favourable effect of the treatment may not have been clearly determined at the time of TES participation (with results from the preceding RCT and/or indeterminate prior studies not available), this raises legitimate ethical issues about the appropriateness of exposing patients to potentially ineffective or only partially effective treatments for additional periods of time.

Challenges of TES

While TES play a valid role, there are clear limitations that should be considered and potential weaknesses in the design and method of analysis that should be addressed.19 TES benefit from the systematic reporting on cumulative drug exposure but have clear limitations in the detection of rare and unexpected events. In addition, selection bias associated with TES populations and lack of generalisability are key factors. These issues are discussed in more detail in the online supplementary material.
This makes interpretation challenging and sometimes unreliable. While guidance for reporting of RCTs \(^{20,21}\) and safety data from biological DMARD registers \(^{22}\) are available, no recommendations for TES in rheumatology have been published to date. \(^{23}\) With this in mind, a task force was created with the principal aim of developing practical recommendations on key aspects of TES on the basis of the European League Against Rheumatism (EULAR) standard operating procedures, \(^{24}\) and thereby a recommended standardised format for future TES data reporting to achieve greater transparency. This manuscript reports the final recommendations as agreed by the task force.

**METHODS**

The task force agreed that a systematic literature review was not indicated for this initiative, as it would merely serve to further establish the lack of consistency in TES and emphasise the need for the development of a standard for future application. \(^{25}\)

The target population for these recommendations was chosen to be rheumatologists, trialists and researchers working in the field of rheumatology, patient organisations and policymakers. The general approach to this project followed the EULAR standardised operating procedures for the elaboration and implementation of evidence-based recommendations. \(^{24}\)

The two task force conveners (MHB and MB) set up a multidisciplinary task force with participants selected based on their field of expertise, knowledge and experience as well as appropriate geographical distribution, primarily across Europe but also North America.

A first meeting of all task force members was convened in January 2011 to primarily define the domains for evaluation. This comprised two breakout sessions, with the task force split into two groups. Each group had a rapporteur who reported the outcome to the whole task force. After a final round of discussion, the task force agreed on the individual items for inclusion in a Delphi exercise. The Delphi method offers a consensus method that is widely used in health service research. \(^{26}\) The two-step Delphi exercise for this initiative was web based, which permitted opinions to be provided and votes on the level of agreement to be cast independently and anonymously. Geographical limitations were also avoided by this approach. It was designed by LS-F and reviewed and modified as indicated by MHB, LC and MB. Details on how the Delphi exercise was formulated, responses were scored and the approach for informing final recommendations was devised can be found in the online supplementary material.

**RESULTS**

**Task force composition**

The multidisciplinary task force comprised 22 participants consisting of 17 rheumatologists, of whom six were clinical epidemiologists and 11 clinical trialists/expert clinicians, two biostatisticians, one fellow and two patient representatives. Participants represented 10 European countries, the USA and Canada.

**Response rate**

Of the 22 invited experts, three could not attend the first meeting (January 2011) but were subsequently apprised of the discussion and participated in the Delphi exercise. One of the patient representatives could not continue participation after the first meeting. Twenty of the 21 participants responded to the first and all 21 responded to the second Delphi exercise.

The two-step Delphi exercise was completed by January 2012, with subsequent analysis and dissemination of draft recommendations in March 2012. Final voting took place in May 2012. However, subsequent steps of involving additional stakeholders (see ‘Results’ section) and a meeting to discuss the recommendations (June 2013) led to a delay in establishing the recommendations for the purposes of submission. The task force approved this final document that included some modifications following the last step. More details on the timelines, responses and involvement of other stakeholders are detailed in the online supplementary material.

**Domains for evaluation**

At the initial meeting, the task force agreed on seven main domains to form the basis of the exercise. These are listed in box 1 with components within each domain that we wished to cover.

**Final results**

Percentage agreement for each recommendation (following the second Delphi exercise) is given. Where appropriate, mean (SD) scores have also been provided. Median (range) scores were also calculated and are included in the online supplementary material.

**Definition of a TES**

- **Study design definition (100% agreement):** A TES is a study that follows all patients beyond a pre-specified trial period whether the trial was (a) a placebo-controlled RCT with the possibility to cross over to usual care or (b) an active comparator RCT.
- **Start of a TES (100% agreement):** Should be stated in the pre-specified protocol with clear justification, and should be at the point of exposure to the experimental drug of interest. For the experimental randomised arm, this will be the start of the original RCT, while for those randomised to placebo/active comparator arm, this point will be on switching to experimental treatment.

**Box 1 The key domains underpinning the Delphi exercise**

1. Definition of a trial extension study (TES)
   - Study design definition
   - Definition of start of TES
   - Duration of TES
   - Patient population of TES
2. Development of a checklist of minimal data items/outcome necessary for a TES
   - Minimal information a TES should collect
   - Elements not amenable to accurate assessment by a TES
   - Safety elements that may be elicited
   - Efficacy
3. Additional data/outcomes
   - Additional legitimate outputs from a TES
4. Method of analysis
5. Method of reporting results
   - Inclusion of a flowchart
   - Detail minimal standards by way of a checklist
   - Frequency and nature of TES
6. Ethics and obtaining consent
7. Over-arching principles
   - Consultation and stakeholder involvement
   - General comments on TES and its reporting
   - Sources of bias and generalisability
**Minimum duration of a TES** (100% agreement): It was agreed by consensus not to define this; nevertheless, the rationale for the length chosen should be stated in the predefined protocol with adequate justification.

**Population for inclusion in a TES** (96% agreement): Should include all patients included in the RCT, with the ability to separately report on patients who are of specific interest, for example, those in remission or with low disease activity.

The minimal information that should be collected and reported by a TES is listed in **Box 2**. Minimum and maximum mean (SD) scores following the first Delphi exercise were 7.2 (2.72) and 9.9 (0.36) (refer to the online supplementary material for individual mean and median scores) with agreement by 100% of the task force in the second Delphi exercise.

The entire group also accepted the following statements relating to the nature of the initial RCT design following the first Delphi exercise:

**Safety and efficacy outcomes**

Evaluation of safety aspects includes several elements, some of which it may not be feasible to capture within certain study designs. The following statements were agreed during the first Delphi exercise (minimum and maximum mean score of 7.0 and 8.4; refer to online supplementary material for individual scores) with 90% accepting all statements in the second round.

Safety

- TES may identify new adverse effects that the original RCT was not able to detect due to greater cumulative drug exposure.
- TES may identify whether the incidence of known adverse effects changes with longer-term drug exposure.
- TES may confirm whether the nature of known adverse effects identified from the RCT changes with longer-term exposure.
- TES are sub-optimal to detect rare safety events because they are not powered for this.
- TES are sub-optimal to detect rare safety events because they include a selected population (responders with likely no previous serious adverse events).

Efficacy

- Greater cumulative exposure to the active drug per patient in a TES might identify additional information on the drug’s efficacy.
- While definitions of relapse are currently not available and require further work, if/when validated, a TES might allow evaluation of relapse including time to relapse and therefore the sustainability of original disease control.

**Additional data/outcomes**

- Economic evaluation of long-term treatment with the active drug may be possible if appropriate measures are recorded in the TES.
- A TES could not accurately evaluate health-related quality of life.

**Method of analysis**

Following the second Delphi exercise, this section required further iterations to refine the initial Delphi statements. These are detailed in **Box 3**. Minimum and maximum scores of agreement were 7.3 and 9.4 (refer to the online supplementary material for individual scores).

**Method of reporting results**

**Inclusion of a flowchart**

- All TES reports should include a flowchart.

This was agreed as a minimal piece of information to accurately illustrate the treatment arms, and changes in treatment and in patient numbers during the course of the study (mean (SD) 9.9 (0.36)).

- In particular, the absolute measure/count should be reported (with/without the percentage).

---

**Box 2 Minimal information to be included in a TES report**

- Progress of subjects at each stage from RCT start to TES completion with:
  - A flow diagram detailing absolute numbers of subjects at each relevant time-point
  - Duration of active treatment
  - Time of last observation
- All drop-outs detailed
- The drop-out rates from each arm during the original RCT and the cross-over groups
- Reason for exclusion from the TES if the patient discontinues the drug
- Reason for cessation of follow-up
- Specification of reasons for cessation of follow-up other than adverse event or ineffectiveness as above, for example, geographical or doctor-related reasons
- Functional status at the time of inclusion in the TES if applicable
- Functional status at last observation if applicable
- Disease activity at the time of inclusion in the TES if applicable
- Disease activity at last observation if applicable
- For those patients entering the TES having achieved low disease activity or remission during the RCT, the sustainability of each disease state should be evaluated and reported
- For those subjects who enter a TES not having achieved remission/acceptable disease activity state following the RCT, the number who achieve this during the TES should be reported to determine whether longer drug exposure has the potential to improve the disease state of such subjects further
- Disease-related co-medication (DMARD, corticosteroid) at each stage from RCT start to TES completion
- Any serious adverse events and outcome related to safety at each stage from RCT start to TES completion DMARD, disease modifying anti-rheumatic drug; RCT, randomised controlled trial; TES, trial extension studies.
The recommendations related to obtaining consent are detailed below; this item in particular required specific input from the patient representative (refer to the online supplementary material for individual scores on additional questions that had means scores of between 6.2–9.4).

- All of the subjects undergoing an RCT should be informed of the importance of long-term surveillance and be given the opportunity of entering in the long-term follow-up (mean (SD) 9.4 (0.85)).
- Subjects should sign a new consent form both for continuation of the drug and for data collection at that time point (mean (SD) 7.6 (2.87)).
- Annual updates for consent are not recommended (mean (SD) 3.7 (4.4)).

Over-arching principles

- The report of a TES should be consistent with and consolidate existing established guidelines including CONSORT20 28 and STROBE29 (mean (SD) score 9.4 (0.85)).
- The report of a TES should be consistent with the ACR/EULAR recommendations on the reporting of clinical trials in RA21 (mean (SD) score 8.9 (1.88)).

General comments on TES and its reporting

All the following statements were accepted by 95% of the group in the second Delphi exercise, agreement with the individual statements having been established as part of the initial Delphi exercise (agreement score out of 10):

- While data linkage is important for long-term observation, access may be difficult as pharmaceutical companies conduct most TES; this may in turn limit the overall benefit of such studies (mean (SD) 7.1 (2.06)).
- TES, by definition, comprise a sub-selected population, not reflective of routine care; hence, even if all patients in an RCT were entered into a TES, such a study is generalisable only to patients with similar disease characteristics (mean (SD) 7.9 (1.76)).
- The absence of a clear null hypothesis may make the definition of comparator groups in a TES difficult (mean (SD) 7.4 (1.74)) and should therefore be stated where appropriate (see table 3 for details on method of data analysis).

Potential sources of bias or lack of generalisability

Several factors were identified as possibly influencing the inclusion of patients in a TES following completion of an RCT, which could introduce sources of bias and lack of generalisability (80% agreement to include all the following statements):

- The requirement of a certain level of response (mean (SD) 7.9 (2.67))
- The stage of the disease of the patient (mean (SD) 7.2 (1.8)).
- The fact that the investigator is remunerated for each patient recruited or that the patients may also receive financial support from the sponsor.

Consent

In a TES, the denominator of a cohort typically decreases over time, which results in the reporting of (artificial) increasing percentages of response rates over time.27 The use of absolute numbers ensures accurate synthesis of the data.

Figure 1 includes a schematic of suggested flowcharts for either placebo-controlled or active comparator RCTs that was accepted by the group (mean (SD) 9.0 (2.06)).

Frequency and nature of reporting outputs from a TES

The following recommendations were made (mean scores between 8.2 and 8.8; refer to the online supplementary material for individual scores):

- Reporting frequency should not be specified for all TES since this depends on the research question.
- However, the protocol of each TES should pre-specify the minimum frequency of reports to be written and the basis for them (purpose, outcomes, length of RCT).
- The efficacy and safety results of a TES should generally be reported together; abstract selection committees and journal editors should carefully consider reporting of efficacy alone before acceptance.
compensation and that the drug is free of charge could be of importance in some health systems (mean (SD) 7.4 (1.7)).

▸ Geographical differences in practice/approach (leading to differences in the number and nature of patients included) (mean (SD) 7.5 (2.45)).

▸ Unwanted heterogeneity from countries where treatment options may be more limited (eg, patients with higher levels of disease activity recruited where otherwise only patients in remission/with low disease activity would be included) (mean (SD) 7.6 (1.45)).

Consultation on recommendations and stakeholder involvement
The Delphi process established whether input from relevant stakeholder organisations, namely, industry, regulatory authorities (Food and Drug Administration (FDA), European Medicines Agency (EMA)) and contract research organisations (CRO) should be sought. In the initial Delphi exercise, 75% voted in favour of some level of industry input, 94% for regulatory authorities and 81% for CRO.

The second Delphi exercise asked for agreement that each of these organisations be included in the initiative:

▸ Industry and regulatory authority input into the final recommendations was recommended, with mean (SD) scores out of 10 of 7.2 (2.48), 8.3 (1.77) and 4.9 (2.85) recorded for the FDA, EMA and CRO, respectively.

Key industry companies that have been associated with new drugs in the RA arena were therefore approached (refer to online supplementary material for details of the companies represented).

DISCUSSION
We present a series of pragmatic recommendations on the design and reporting of TES in rheumatological conditions (mainly inflammatory arthritis, although the basic principles are generally applicable), based on a high degree of expert consensus. Our EULAR task force comprised a group of experts encompassing a range of expertise including clinical trialists, clinicians experienced in RA treatment, and clinical epidemiologists as well as patient representatives. A wide range of countries...
and health systems were represented, albeit with some omissions (eg, absence of individuals from Asia), although the opportunity to evaluate these recommendations in the wider community in the future should highlight any differing perspectives. With a generally accepted methodology for prospective observational studies, we felt an additional systematic review was not necessary and decided to use our expert opinion to formulate guidance for TES. These recommendations complement those established for clinical trials and registries.

Central to the recommendations was the principle that a TES report should focus on cumulative outcome analysis, maintaining the original trial groups to avoid reporting of only the sub-selected patient group that proceeds to the TES, and thereby achieve better generalisability of results. Furthermore, the task force was clear that absolute numbers and not just percentage response rates should be reported. To facilitate this, we recommend a flow diagram detailing absolute numbers of subjects at each relevant time point, with clear illustration of drop-outs and the reason for cessation and/or exclusion at each relevant stage. While it was agreed that a TES might elaborate on the incidence and nature of adverse events over time, they are not designed to capture rare safety signals. TES reports may also have the potential to inform on the durability of response and the dynamics of achieving pre-determined targets of treatment (low disease activity and remission). It was agreed that any analysis should be prespecified in the protocol but should always include an intention-to-treat in addition to a completer approach. We acknowledge there are elements that may in particular be the subject of further discussion in the wider community, for example, the issue of split reporting. While the task force discouraged this, each case should be considered individually as there may be instances when there is utility in this approach to ensure relevant data that is of interest is disseminated within the public domain.

The recommendations were actively commented on by several industry companies (see the ‘Consultation of recommendations and stakeholder involvement’ section) and include their specific feedback (which has been indicated directly in the results where appropriate in the online supplementary material) and as such, gained the approval of the participating stakeholders. While EMA representation did not suggest changes to the recommendations, it acknowledged the importance of standardisation. The interaction also highlighted how regulatory expectations may drive the industry approach on whether and how TES should be undertaken.

While we acknowledge that the working group was perhaps relatively small for a consensus exercise, following dissemination of these recommendations, we would anticipate a subsequent exercise to capture how they have been received in the wider rheumatology, trial and industry communities. In future, it will be important for journal reviewers and editors to measure future TES reports against the standard set by these recommendations. The future research agenda will include a systematic review of forthcoming TES to evaluate how well this document is utilised, with further refinement based on the nature of outcomes observed. In addition, regulatory agencies may wish to consider the recommendations and associated issues and how these may influence their expectations from industry. This initiative and the interactive session at EULAR, Madrid 2013 with relevant stakeholders will hopefully be a springboard for further action (the outcome of the EULAR meeting is summarised in the online supplementary material).

In summary, there is a clear unmet need for a reliable approach to the reporting of TES to maximise our understanding of drug effects in chronic conditions. This initiative, its principles and resulting recommendations apply to TES for any drug in RA as well as for drugs used to treat other chronic rheumatological conditions. This document provides much needed first recommendations to ensure a transparent and standardised approach to the reporting of future TES.

**Author affiliations**

1. Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, UK
2. NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds Teaching Hospitals NHS Trust, Leeds, UK
3. Rheumatology Department, Hospital Universitario de Guadalajara, Guadalajara, Spain
4. Institute for Musculoskeletal Health, Madrid, Spain
5. Department of Internal Medicine 3, Division of Rheumatology, Medical University of Vienna, Vienna, Austria
6. Department of Rheumatology, Musculoskeletal Statistics Unit, The Parker Institute, Copenhagen University Hospital, Bispebjerg and Frederiksberg, Copenhagen, Denmark
7. Department of Rheumatology, Lapeyronie Hospital, Montpellier I University of Montpellier, France
8. Division of Rheumatology, Catholic University of the Sacred Heart, Rome, Italy
9. Université de Lorraine, Université Paris Descartes, Nancy, France
10. Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway
11. Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands
12. Institute of Rheumatology and Clinic of Rheumatology, Charles University, Prague, Czech Republic
13. Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham School of Medicine, Birmingham, Alabama, USA
14. 2nd Department of Medicine, Center for Rheumatic Diseases, Hietzing Hospital, Vienna, Austria
15. Arthritis Research UK Centre for Epidemiology, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
16. NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Trust, Manchester, UK
17. Leiden University Medical Center, Department of Rheumatology, Leiden, The Netherlands
18. EULAR Patient Research Partner, Ede, The Netherlands
19. Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada
21. Department of Rheumatology, University Hospitals Leuven, Leuven, Belgium
22. German Rheumatism Research Centre, Berlin, Germany
23. Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

**Acknowledgements** We would like to thank the European League Against Rheumatism (EULAR) for supporting and providing the funds for this task force initiative.

**Contributors** MHB, LS-F, LC and MB contributed substantially to the design, implementation and data collection of the Delphi exercises. LS-F analysed the DELPHI data. MHB, LS-F, LC and MB reviewed the DELPHI data analysis before dissemination to the task force. MHB wrote the paper and the supplementary materials. All authors discussed the summaries presented in the Delphi exercises, results and implications and commented on the manuscript at all stages.

**Funding** MHB was supported by a National Institute of Health and Research (NIHR) Clinician Scientist Award. RC is based at the Musculoskeletal Statistics Unit, The Parker Institute, which is supported by grants from the Oak Foundation.

**Competing interests** None.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

**REFERENCES**


ON LINE SUPPLEMENTARY MATERIAL

Challenges of TES

High potential for selection bias may be introduced as would be observed if for example, all patients move onto one arm (open label drug), which is a frequent approach. Even if randomisation is maintained however (with an active comparator arm), loss of generalisability (and therefore external validity) is introduced – with losses to follow-up and patients that respond poorly to the experimental agent as well as those that experience major adverse events are often withdrawn from RCTs. RCT participants who have suboptimal response or experience mild intolerance may complete the RCT but are less likely to continue participation in a TES. Analysis of a TES does not usually include such patients but solely focuses on those who enter the open-label extension period. By not taking this into account in the method of analysis, researchers tacitly assume that patients not entering a TES have outcomes similar to the TES patients. In other words, this resembles a ‘study completers’ analysis, with the inappropriate assumption that the outcome of subjects not entering (or completing) the TES is ‘missing at random’[2]. The end result is usually a favourable, but biased picture of the long-term assessment of benefit and harm of the study drug. In addition, the description of the population comprising and progressing in a TES report is frequently unclear and incomplete, making it difficult to be able to inform decisions about practice[3]. This includes combining of patients newly starting the drug (having previously been randomised to the comparator arm) with those who have already been on it for the duration of the RCT.

Methods

The first Delphi questionnaire (see appendix A) was accompanied by an explanation of the purpose of the exercise and sent to all the members of the task force who were asked to respond within 2 months. It included 21 items based on the 7 main domains for evaluation. Each item included a variable number of questions; some required a “yes” or “no” answer and others asked for a level of agreement of a statement on a scale from 0 to 10 to be chosen. All items had space for additional comments. (See appendix A for the questionnaire used.) We calculated the
proportion of respondents who answered yes and no to each question included in the questionnaire. The steering group (MHB, MB, LC) arbitrarily a priori decided that group acceptance would be defined as items in which ≥70% respondents responded similarly (for questions requiring ‘yes or no’ answers). Statements requiring marking of a level of agreement were accepted if a mean score of ≥ 7/10 was recorded. Statements were rejected if a mean score of < 7 was noted and/or if more than 4 responses of the task force of 22 under 5 were recorded.

The initial set of items was amended based on the analysis and comments of the first round. A second Delphi was subsequently developed (by LS-F and MHB/LC/MB as before) that included 26 items for rating (see appendix B). This second questionnaire and the results from the first questionnaire were sent by e-mail to the task force members. Participants were asked to respond, to a second web-based Delphi survey within 6 weeks and after 4 weeks, e-mail reminders were sent to any non-responders.

Agreement of items formed the basis of the recommendations. A final and third round of discussion was undertaken electronically to modify any of the statements following which a voting round was undertaken to determine whether consensus had been reached. Finally, the task force also agreed upon the formulation of a research agenda.

Timelines

*Stage 1 Delphi exercise*
Stage 1 of the Delphi survey (appendix A) was mailed to all participants in April 2011. All responses were received by June 2011. This stage comprised 20 items, which included formally accepting the domains for evaluation summarised above. Opportunity for comments to raise additional points for consideration was also included at the end of each item.
Stage 2 Delphi exercise

A second Delphi was developed following on from the analysis and comments of the first Delphi survey (appendix B). This was sent to all participants in August 2011. All responses were received by December 2011 and reported to the Task Force in January 2012. These were analysed and a further round of modifications and amendments were made to 2 particular items following electronic discussion. The final set of recommendations was produced and a request for voting on these was sent in March 2012. All 21 task force participants voted by May 2012, accepting the recommendations.

The task force agreed to seek formal industry and regulatory body input before submitting these recommendations (triggered by agreement in one of the Delphi questions—see results). Obtaining agreement from several industry companies as well as European Medical Agency took some time; with the annual EULAR congress then seen as an appropriate setting to convene a meeting. As will be noted, some modifications were subsequently made but all responses were not collated until Autumn 2013. All members of the task force re-approved the recommendations by way of confirmation prior to document submission.

Results

Definition of a TES

The initial definition, which achieved the highest mean (SD) score 7.7 (2.55); median (range) 9 (1-10), was selected and subsequently modified based on additional comments made by the participants. In the second Delphi survey, the following revised definition had final 100% agreement, “A TES is a study that follows all patients beyond a pre-specified trial period whether the trial was a) a placebo-controlled RCT with the possibility to cross-over to open-label experimental drug or b) a placebo-controlled RCT with the possibility to cross-over to usual care or c) an active comparator trial.”

Industry input highlighted that this definition may exclude studies where cross-over to other treatments are included and was therefore modified to the following final
definition of a TES: “A TES is a study that follows all patients beyond a pre-specified trial period whether the trial was a) a placebo controlled RCT with or without the possibility to cross-over to open-label experimental drug or b) an active comparator trial.”

Definition of the start of a TES
The starting point of a TES should be stated in the pre-specified protocol with clear justification; and should be at the point of exposure to the experimental drug of interest (100% acceptance). For the experimental randomised arm this will be start of the original RCT; whilst for those randomised to placebo/active comparator arm, this point will be on switching to experimental treatment (during RCT or at start of TES, see later and figure 1).

Minimum duration of a TES:
The committee could not reach consensus on whether or not a minimum length of a TES should be defined (68% agreement to define), as this would be determined by the research question. The task force therefore agreed not to define a minimum duration; nevertheless, the rationale for the length chosen should be stated in the pre-defined protocol with adequate justification (100% agreement).

Population for inclusion in a TES:
All but one of respondents agreed that the population of TES should not be stipulated in guidelines, as this would be determined by the individual research question. Ideally, however, it should include all patients included in the RCT, with the ability to separately report on patients who are of specific interest, for example, those in remission or low disease activity.

Minimal data items/outcomes
Table 2 from the main manuscript is included below, with the addition of individual mean/median agreement scores.
<table>
<thead>
<tr>
<th>Nature of information</th>
<th>Agreement, mean (SD) score, 1-10</th>
<th>Agreement, median (range) score, 1-10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Progression from RCT to TES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progress of subjects at each stage from RCT ^start to TES* completion with:</td>
<td>8.7 (2.2)</td>
<td>10 (2-10)</td>
</tr>
<tr>
<td>A flow diagram detailing <strong>absolute</strong> numbers of subjects at each relevant time-point</td>
<td>9.9 (0.36)</td>
<td>10 (9-10)</td>
</tr>
<tr>
<td>Duration of active treatment</td>
<td>9.5 (0.65)</td>
<td>10 (8-10)</td>
</tr>
<tr>
<td>Time of last observation</td>
<td>9.5 (0.94)</td>
<td>10 (7-10)</td>
</tr>
<tr>
<td><strong>Patient drop-outs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All drop-outs detailed</td>
<td>9.1 (1.46)</td>
<td>10 (5-10)</td>
</tr>
<tr>
<td>The drop-out rates from each arm during the original RCT and the cross-over groups</td>
<td>9.3 (1.07)</td>
<td>10 (7-10)</td>
</tr>
<tr>
<td>Reason for exclusion from the TES if the patient discontinues the drug</td>
<td>9.5 (1.02)</td>
<td>10 (7-10)</td>
</tr>
<tr>
<td>Reason for cessation of follow-up</td>
<td>9.4 (1.0)</td>
<td>10 (7-10)</td>
</tr>
<tr>
<td>Specification of reasons for cessation of follow up other than adverse event or</td>
<td>8.7 (1.82)</td>
<td>9.5 (4-10)</td>
</tr>
<tr>
<td>inefficacy as above, e.g. geographical or doctor related reasons</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional status at the time of inclusion in the TES if applicable</td>
<td>8.8 (1.67)</td>
<td>9.5 (4-10)</td>
</tr>
<tr>
<td>Functional status at last observation if applicable</td>
<td>8.3 (1.73)</td>
<td>8 (4-10)</td>
</tr>
<tr>
<td>Disease activity at the time of inclusion in the TES if applicable</td>
<td>9.3 (0.91)</td>
<td>10 (8-10)</td>
</tr>
<tr>
<td>Disease activity at last observation if applicable</td>
<td>9.4 (0.85)</td>
<td>10 (8-10)</td>
</tr>
</tbody>
</table>
For those patients entering the TES having achieved low disease activity or remission during the RCT, the sustainability of such disease states should be evaluated and made available.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.6 (1.22)</td>
<td>8 (7-10)</td>
</tr>
</tbody>
</table>

For those subjects that enter a TES not having achieved remission/acceptable disease activity state following the RCT, the number that achieve this during the TES should be reported – to determine whether longer drug exposure has the potential to improve disease state of such subjects further.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.3 (1.59)</td>
<td>8 (5-10)</td>
</tr>
</tbody>
</table>

### Treatment

The disease –related co-medication (DMARD#, corticosteroid) at each stage from RCT start to TES completion.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.2 (2.72)</td>
<td>8 (0-10)</td>
</tr>
</tbody>
</table>

### Safety

The serious adverse events and any outcome related to safety at each stage from RCT start to TES completion.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.3 (2.08)</td>
<td>8.5 (2-10)</td>
</tr>
</tbody>
</table>

### Safety

- TES may identify new adverse effects that the original RCT was not able to detect due to greater cumulative drug exposure; mean (SD) 8.4 (1.65); median (range) 9 (4-10).
- TES may identify whether the *incidence* of known adverse effects changes with longer-term drug exposure; mean (SD) 7.5 (1.61); median (range) 7.5 (4-10).
• TES may confirm whether the nature of known adverse effects identified from the RCT changes with longer-term exposure; mean (SD) 7.6 (1.5); median (range 7.5 (5-10).
• TES are sub-optimal to detect rare safety events because they are not powered for this; mean (SD) 7.6 (2.71); median (range) 8 (0-10).
• TES are sub-optimal to detect rare safety events because they include a selected population (responders with likely no previous serious adverse events); mean (SD) 7.0 (2.75); median (range) 8 (1-10).

**Efficacy:** The task force agreed that the greater cumulative exposure of the active drug per patient in a TES might identify additional information on the drug’s efficacy; mean (SD) 7.0 (2.14); median (range) 7.5 (3-10). Whilst definitions of relapse are currently not available and requires working on, if/when validated, a TES might allow evaluation of relapse including time to relapse and therefore sustainability of original disease control; mean (SD) 7.8 (1.42); median (range) 8 (5-10).

**Additional data/outcomes**
Possible additional outputs to safety and efficacy were explored. Economic evaluation of long-term treatment with the active drug may be possible if appropriate measures are recorded in the TES; mean (SD) 7.2 (2.33); median (range) 7.5 (0-10). The committee did not accept that a TES could accurately evaluate health-related quality of life; mean (SD) 6.6 (2.35); median (range) 7 (0-10), risk-benefit ratio and therefore overall advantage of the drug; mean (SD) 6.2 (2.96); median (range) 7 (0-10), or compliance; mean (SD) 6.1 (3.09); median (range) 7 (0-10).

Method of data analysis
Table 3 from the main manuscript is included below, with the addition of individual mean/median agreement score

<table>
<thead>
<tr>
<th>Statement</th>
<th>Agreement, mean (SD) score</th>
<th>Agreement, median (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The null hypothesis should be stated at the start where appropriate</td>
<td>7.9 (2.16)</td>
<td>8.5 (4-10)</td>
</tr>
<tr>
<td>Multiple comparisons should be taken into account when determining the level of statistical significance</td>
<td>8.1 (1.7)</td>
<td>8.5 (5-10)</td>
</tr>
<tr>
<td>The null hypothesis should take account of the results of the original RCT*. Depending on the research question, the results of a RCT should be accommodated in the TES*</td>
<td>7.5 (2.19)</td>
<td>8 (1-10)</td>
</tr>
<tr>
<td>The report should comment on cumulative outcome analysis (beneficial and adverse events) maintaining the original trial groups i.e. from RCT start, not TES start to avoid reporting of only the sub-selected patient group that proceeds onto the TES</td>
<td>8.6 (1.54)</td>
<td>9 (5-10)</td>
</tr>
<tr>
<td>The selection bias associated with a TES population means meaningful non-inferiority/superiority analysis would not be reliable. The report should focus on how data for sustained effect from the start to the end of TES period, within a single group or the difference between groups was analysed and whether there was any suggestion of increased effect (although this could not be subject to formal statistical testing).</td>
<td>9.4 (0.84)</td>
<td>10 (8-10)</td>
</tr>
<tr>
<td>The plan for subjects that drop out of a TES</td>
<td>8.9 (0.91)</td>
<td>9 (8-10)</td>
</tr>
</tbody>
</table>
should be specified to demonstrate sustained effect from the start to end of TES period. With reducing number of participants (the denominator), the proportion responding will artificially increase if/when the number of patients (numerator) responding stays the same.

| The analysis should include survival/retention rates on therapy explicitly reporting the number of patients at each milestone with reasons for change detailed. | 8.9 (1.29) | 9.5 (7-10) |
| A plan on how to analyse this should be included with both intent-to-treat (ITT) (denominator as original number entering RCT) and completer (those entering TES only) population analyses reported. A completer analysis should always be reported together with an ITT analysis. | 9.3 (0.99) | 10 (7-10) |
| The repeated measures analysis of the data from a TES in rheumatology should include the area under the curve of absolute disease activity (i.e. not dichotomous response/change) preferentially expressed as a score (e.g., DAS, SDAI, etc.) | 7.3 (2.55) | 8 (1-10) |
| A TES should preferably include hard endpoints (e.g. death, work disability, joint replacement surgery, hospital admission) from TES +/- linkages with other data sources | 8.6 (1.28) | 8 (7-10) |
Frequency and nature of split reporting
The committee agreed that reporting frequency should not be specified for all TES since this depends on the research question. In addition, industry representatives highlighted the fact that data cuts and reports may be undertaken in response to regulatory considerations that may not be foreseen. It was agreed that the protocol of each TES should pre-specify the minimum frequency of reports to be written and the basis for them (purpose, outcomes, length of RCT); mean (SD) 8.2 (2.28); median (range) 9 (1-10). Regarding the nature of reports, it was agreed that the results of efficacy and safety of a TES should be reported together when all patients have reached the specified time point as applicable; [mean (SD) 8.8 (1.36); median (range) 9 (5-10). The credibility of split reporting (for example, one abstract on efficacy, one on safety, one on quality of life outcomes) is questionable and should be discouraged by abstract selection committees and journal editors; mean (SD) 8.7 (1.75); median (range) 10 (5-10). It was acknowledged following stakeholder feedback however that abstract length limitations can be significant, and that certain complex safety communications focusing on specific events of interest would suffer from requirement to include efficacy information. The above recommendation was modified to, “the results of efficacy and safety of a TES should generally be reported together; abstract selection committees and journal editors should carefully consider reporting of efficacy alone before acceptance”.

Consent:
The group accepted that all of the subjects undergoing a RCT should be informed of the importance of long-term surveillance and be given the opportunity of entering in the long-term follow-up; mean (SD) 9.4 (0.85); median (range 10 (8-10). Subjects included in a TES should sign a new informed consent form for continuation of data collection; mean (SD) 7.6 (2.87); median (range) 8.5 (1-10). The patient representative emphasised the importance of the patient knowing when they have transitioned from the RCT to TES; hence would be in favour of them signing a new consent form both for continuation of the drug and for data collection at that time point; mean (SD) 7.6 (2.87); median (range) 8.5 (1-10). Although it was agreed that
annually updating the consent of patients included in a TES was not necessary; mean (SD) 3.7 (4.4); median (range) 1.5 (0-10); particularly since each additional consent runs the risk of additional drop-out, the statement in the second survey that this need not occur only achieved a mean (SD) score of 6.2 (3.79); however, median (range) score of 8 (0-10) was noted. Nevertheless, the patient representative advised we should not recommend the need for consent to be annually updated.

These comprised: AbbVie, Bristol-Myers Squibb, Merck-Sharp & Dohme, Pfizer, Roche, UCB. Open discussion was subsequently held at a EULAR annual congress 2013 session (see discussion) where Dr Cesar de la Fuente Honrubia (Spanish Agency of Medicines and Medical Devices, European Medicines Agency) and Professor Paul Peter-Tak (Senior Vice President/Head, ImmunoInflammation, Glaxo SmithKline) provided regulatory and industry perspectives respectively. The following companies attended EULAR +/- actively participated in the discussion of the recommendations and provided feedback to the final document: AbbVie, Bristol-Myers Squibb, Merck-Sharp & Dohme, Pfizer, Roche.

**Discussion**

*Outcome following interactive session at EULAR, Madrid 2013: Consensus on steps going forward*

Whilst the participating stakeholders approved the recommendations, a fundamental question was raised for future consideration - should industry undertake open-label extension studies in the first place? Central to this is establishing what the objectives of a TES are and whether the TES is the most appropriate method and best use of resources. Similar issues to those detailed earlier questioning the validity of TES and the evidence that switching from placebo to active improves disease control were raised. The notion that it is important for those who have benefitted to stay on drug perhaps contradicts the justification for a study that assumes equipoise between the 2 treatment options. It may be more appropriate to offer the trial participant that benefitted to continue allocated treatment in a blinded fashion – thereby preserving the rigour of efficacy and safety data. The expectation of regulatory authorities was highlighted as a crucial factor driving the conduct of such studies; a large proportion of industry-sponsored TES are
conducted as a result of requests made by the regulatory authorities during the registration process. This is not only to acquire longer-term efficacy data but often mainly to address specific safety concerns.

The meeting closed with a general agreement to pursue this field further; with the acknowledgment that any future discussion and consideration amongst the clinical academic community for change will rely on the engagement of both industry and regulatory authorities. This report is therefore seen as an initial phase of a wider initiative and a springboard for further development.
References

Dear Taskforce Member,

Please find below a survey on the design and reporting of long-term extension (LTE) studies as part of your task as a panel member.

The two main objectives of the survey are:

1. To define the items that should be included in a recommendation document.
2. To establish the level of agreement of statements in the document.

Your considered opinion is crucial to the successful outcome of this project.

Please read each section carefully before answering. You will be asked to answer specific questions (with graded response). There will also be the opportunity to include free text and comment.

We would like to emphasise that there are no absolute wrong or right answers. At the end of the survey, you will have available an open-ended box for comments.

The survey is reasonably detailed and is likely to take 30 minutes to complete. Please make sure that you do not leave the survey half completed.

If you do not feel sufficiently familiar or confident to answer any of the questions, please leave it blank.

Maya Buch, Marteen Boers, Lucia Silva, and Loreto Carmona
Long-term extension (LTE) studies in Rheumatology

Aspects to be dealt with in the taskforce

1. How important do you feel it is for opinions of each of the following stakeholders to be included in the taskforce final document?

Please choose the most appropriate form of involvement for each individual

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>None</th>
<th>Offer to provide comment on the final document</th>
<th>Offer to provide comment on one or more drafts</th>
<th>Participation in the taskforce</th>
</tr>
</thead>
<tbody>
<tr>
<td>An industry representative</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An administration (FDA, EMA) representative</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A contract research organisation representative</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (please specify)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. How important do you feel it is for each of the following items to be included in the taskforce final document?

Please grade the level of priority or importance (0 = very low; 10 = extremely important)

<table>
<thead>
<tr>
<th>Item</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A reference to the STROBE guidelines for reporting observational studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A definition of LTE studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A paragraph on the strengths and limitations of LTE Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A checklist with minimal data the study should collect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendations on how to build a flow chart of the study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendations on how to analyse the data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendations on how to report the results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethical comments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Are there other items or aspects that the taskforce / document should address?

[Input field for additional comments]
Long-term extension (LTE) studies in Rheumatology

Definition of a long-term extension (LTE) study

The definition of a LTE study is in itself not clear and remains to be established. Several areas warrant consideration when determining the definition, such as design, length of study and the population that should be studied.

For the following statements on the definition of a LTE study, please grade your level of agreement (0 = none to 10 = maximum).

3. STUDY DESIGN DEFINITION

Please rate your level of agreement (0 = none; 10 = maximum)
(Options may overlap)

"A long-term extension study is any study that..."

<table>
<thead>
<tr>
<th>Statement</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>...follows patients beyond the pre-specified RCT period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...follows patients beyond the pre-specified RCT period, on condition that all patients entering the original trial are accounted for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...follows a placebo-controlled study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...follows an active-comparator study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...follows a placebo-controlled study in which patients start the experimental treatment either as a result of randomised allocation or crossover after placebo treatment; and only patients on experimental treatment are followed beyond the pre-specified trial period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...follows all patients beyond the pre-specified trial period whether the trial was: A) a placebo-controlled RCT with the possibility to cross-over to open label drug; OR B) a placebo-controlled RCT with the possibility to cross-over to usual care; OR C) an active comparator trial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Can you think of a better definition of the design of a LTE study? If so, please include free text in the box below

4. MINIMUM LENGTH of LTE STUDY

Towards defining the minimum length of a long-term extension study, which of the following 2 options would you choose as the starting point of a LTE study?

- [ ] Start is on completion of the pre-specified RCT period
- [ ] Start is at the point of cross-over or switch of therapy as allowed in the RCT
5. Using your preferred starting point, how many YEARS should a LTE study last?

- [ ] up to 12 months
- [ ] 13-24 months
- [ ] 25-36 months
- [ ] More than 3 years
- [ ] The minimum length of an LTE need not be defined

Combining questions 4 and 5, can you think of a better definition of the length of a LTE study? If so, please include as free text in the box below:
6. PATIENT POPULATION

What patient population should a LTE study inform us on and therefore be included in such a study?

Please rate your level of agreement with the following statements (0 = none; 10 = maximum)

<table>
<thead>
<tr>
<th>Statement</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Should include all patients initially included in the RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Should include only those patients who achieved remission during the RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Should include only those patients who achieved remission or a low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disease-activity (LDA) state during the RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May include patients with any level of disease activity on completion of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the RCT (remission, LDA or moderate-high disease activity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Can you think of a better definition of the patient population that should be included in a LTE study? If so, please include as free text in the box below
Long-term extension (LTE) studies in Rheumatology

Potential problems of LTE studies

LTE studies raise other issues and concerns that may also be worth including in the final document.

7. For each of the following statements illustrating POTENTIAL CHALLENGES of a LTE study, please grade your level of agreement to mention in the final document (0 = none to 10 = maximum)

<table>
<thead>
<tr>
<th>Statement</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare safety events may not be detected for reasons other than the length of observation</td>
<td></td>
</tr>
<tr>
<td>Since LTE are largely supported by pharmaceutical companies, the potentially limited access to data linkages (which are important for long-term observations) may further question the overall benefits of such studies</td>
<td></td>
</tr>
<tr>
<td>LTE studies may be extremely costly making them difficult to justify</td>
<td></td>
</tr>
<tr>
<td>Even if all patients in an RCT were entered into an LTE, such a study is generalizable only to patients with similar disease characteristics; many trial populations do not reflect patients seen in routine care</td>
<td></td>
</tr>
<tr>
<td>The definition of comparator groups in a LTE study may be difficult because of the absence of a clear null hypothesis</td>
<td></td>
</tr>
<tr>
<td>Unwanted heterogeneity may result by wishing to accomodate countries where treatment options may be more limited (e.g. allowing patients with higher levels of disease activity to be recruited where otherwise inclusion is of patients in remission/LDA state only)</td>
<td></td>
</tr>
</tbody>
</table>

If you believe there may be other challenges with a LTE study that have not been mentioned and would be worth including in the document, please list them below.

---

8. The following statements illustrate potential SOURCES OF BIAS OR LACK OF GENERALISABILITY of a LTE study.

Please grade your level of agreement to mention in the final document (0 = none to 10 = maximum)

The inclusion of patients in a LTE study following completion of a RCT...

<table>
<thead>
<tr>
<th>Statement</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>...usually requires a certain level of response</td>
<td></td>
</tr>
<tr>
<td>...may be influenced by the fact that the investigator is remunerated for each patient recruited</td>
<td></td>
</tr>
<tr>
<td>...may be influenced by geographical differences in practice / approach (leading to differences in number and nature of patients included)</td>
<td></td>
</tr>
<tr>
<td>...may be influenced by the stage of the disease of the patient</td>
<td></td>
</tr>
</tbody>
</table>

If you believe there may be other sources of bias in a LTE study that have not been mentioned and would be worth including in the document, please list them below.

---
The investigative objective of LTE studies may not be clear to many observers.

For the following statements on justifiable outcomes of a LTE study, please grade your level of agreement (0 = none to 10 = maximum).

### 9. SAFETY

**A long-term extension study...**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Score Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>...may identify new adverse effects that the original RCT was not able to</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>detect due to greater cumulative exposure to the drug</td>
<td></td>
</tr>
<tr>
<td>...is not powered for rare adverse events and so should not be relied upon</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>to detect safety signals</td>
<td></td>
</tr>
<tr>
<td>...may identify whether the incidence of known adverse effects show</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>change with longer-term drug exposure (e.g. infection risk)</td>
<td></td>
</tr>
<tr>
<td>...may confirm whether the nature of known adverse effects identified from</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>the RCT changes with longer-term exposure (e.g. infection risk)</td>
<td></td>
</tr>
<tr>
<td>...is inappropriate to detect rare safety events due to the inclusion of a</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>selected population (responders with likely no previous adverse events)</td>
<td></td>
</tr>
</tbody>
</table>

### 10. EFFICACY

<table>
<thead>
<tr>
<th>Statement</th>
<th>Score Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>The greater cumulative exposure of the active drug in an LTE study</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>may identify additional information on the drug’s efficacy</td>
<td></td>
</tr>
<tr>
<td>A LTE study may allow evaluation of relapse including time to relapse (for</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>example, patients entering the LTE study having achieved an acceptable</td>
<td></td>
</tr>
<tr>
<td>state, such as LDA or remission, on the active drug can be followed to</td>
<td></td>
</tr>
<tr>
<td>assess sustainability</td>
<td></td>
</tr>
</tbody>
</table>

### 11. ADDITIONAL OUTPUTS

Besides efficacy and safety, an LTE may allow...

<table>
<thead>
<tr>
<th>Statement</th>
<th>Score Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>...economic evaluation of long-term treatment with the active drug</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>...better Health-related quality of life (QoL) analysis</td>
<td></td>
</tr>
<tr>
<td>...better evaluation of the risk-benefit ratio and therefore overall</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>advantage of the drug</td>
<td></td>
</tr>
</tbody>
</table>

### 12. In addition to the list above, what other outcomes do you think a LTE study may be able to identify?

Please include your comments in the box below
Long-term extension (LTE) studies in Rheumatology

Minimum amount of information a LTE study should include

This next section focuses on specific recommendations related to the conduct of LTE studies.

13. The following items are proposals for the minimal information a LTE study should collect.

Please grade your level of agreement (0 = none to 10 = maximum)

<table>
<thead>
<tr>
<th>Item</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>The time of last observation</td>
<td></td>
</tr>
<tr>
<td>The functional status at the time of inclusion in the LTE</td>
<td></td>
</tr>
<tr>
<td>The functional status at last observation</td>
<td></td>
</tr>
<tr>
<td>The disease activity at the time of inclusion in the LTE</td>
<td></td>
</tr>
<tr>
<td>The disease activity at last observation</td>
<td></td>
</tr>
<tr>
<td>The reason for exclusion from the LTE study if the patient discontinues the drug</td>
<td></td>
</tr>
<tr>
<td>The reason for cessation of follow-up</td>
<td></td>
</tr>
<tr>
<td>Specification of reasons for cessation of follow up other than adverse event or inefficacy, e.g. geographical or doctor related reasons</td>
<td></td>
</tr>
<tr>
<td>The progress at each stage from RCT start to LTE completion</td>
<td></td>
</tr>
<tr>
<td>The duration of active treatment</td>
<td></td>
</tr>
</tbody>
</table>

In addition to the list above, what other data do you think should be collected as a minimum requirement? Please provide comments in the box below:

14. Do you think the minimum data requirements should be different depending on whether the LTE study has followed a placebo-controlled or an active comparator trial?

- Yes
- No

If you answered 'yes', please describe what should be different.

15. FOLLOW-UP

Please grade your agreement with the following statement (0=none, 10=maximum)

An LTE study that follows an active-comparator RCT should follow all randomised patients for the same period of time (not only patients on the experimental treatment)
### Analysis and interpretation

#### 16. Regarding specific recommendations on the analysis of LTE studies, please rate your level of agreement to the statements below (0 = none to 10 = maximum).

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>The null hypothesis should be stated at the start RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The null hypothesis must be related to the results achieved in the original RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple comparisons should be taken into account when determining the level of statistical significance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The report should include details on how data for sustained effect was analysed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The plan for subjects that drop out of a LTE study should be specified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In a LTE study, the planned analysis of data to evaluate for sustained effect should be non-inferiority in nature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The analysis of the data from a LTE study in rheumatoid arthritis (RA) should include the area under the curve of absolute disease activity (i.e. not response/change) preferentially expressed as a score (DAS, SDAI, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The analysis should include survival / retention rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The analysis should include a pooled analysis from the original trial groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An LTE study should preferably include hard endpoints (e.g. death, work disability, joint replacement surgery, hospital admission) from linkages with other data sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The analysis of the data from a LTE study should take into account the dropouts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please provide any additional comments/points that need to be considered related to the analysis and interpretation of LTE study data in the box below.
Patients randomised to either group at the RCT may leave the trial early. Patients on the active arm may continue receiving the drug until the end of the RCT or discontinue it before ending. Patients on the placebo group may continue on placebo until the end of the RCT or start on the experimental drug before ending. At the start of the extension period two subgroups will be receiving the experimental drug and those in placebo may also commence active drug if they enter the LTE phase.

17. Please grade your level of agreement based on the flowchart above.

<table>
<thead>
<tr>
<th>All LTE reports must include a flow-chart on the progress of patients included</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade your agreement with the flowchart above</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please provide additional comments on the flowchart in the space provided below:
18. Regarding specific recommendations on how to report the results of LTE studies, please rate your level of agreement (0 = none) to (10 = maximum) for the following statements:

(Some of these statements overlap with those made earlier, but we are now asking your opinion on what should always appear in the report)

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>The taskforce should develop and detail minimal standards that should be included when reporting a LTE study by means of a checklist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The report of a LTE study should be consistent with and consolidate existing established guidelines including CONSORT and STROBE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The report of a LTE study should be consistent with the ACR/EULAR recommendations on the reporting of clinical trials in RA (Aletaha D, et al 2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A report of a LTE study should include a flow diagram detailing numbers at each relevant time-point</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTE studies should provide the same information for all randomised patients in terms of quantity and quality, irrespective of the nature of prior RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For those patients entering the LTE study having achieved LDA or remission during the RCT, the sustainability of such disease states should be evaluated and made available</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For those subjects that enter a LTE study not having achieved remission/acceptable disease activity state following the RCT, the number that achieve this during the LTE study should be reported – to determine whether longer drug exposure has the potential to improve disease state of such subjects further</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The drop-out rates from each arm during the original RCT and the cross-over groups should be available</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All drop-outs should be detailed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please provide any additional comments/points you believe may be relevant to the reporting of LTE
**Long-term extension (LTE) studies in Rheumatology**

19. Regarding the nature and frequency of reports from LTE studies, please rate your level of agreement (0 = none) to (10 = maximum) for the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>The taskforce document should comment on the frequency of the reports of LTE studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The taskforce document should comment on the nature of reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The results of a LTE study should be reported every year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The results of a LTE study should be reported every two years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The results of a LTE study should be reported every three years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The results of a LTE study should be reported annually to a maximum of 5 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The results of efficacy and safety of a LTE study should be reported together</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separate reports for efficacy and safety results of a LTE study should be written</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The credibility of split reporting (e.g. one abstract on efficacy, one on safety, one on QoL etc) – is questionable and should be discouraged by abstract selection committees and journal editors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please provide any additional comments/points you believe may be relevant to the frequency or nature of LTE reports.
20. Regarding specific recommendations on the ethical issues involving LTE studies, please rate your level of agreement (0 = none) to (10 = maximum) for the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>All of the subjects undergoing a RCT should be informed of the importance of long-term surveillance and be given the opportunity of entering in the long-term follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The subjects included in a LTE study should sign a new informed consent form (different from the one for the RCT) for continuation of data collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The subjects included in a LTE study should sign a new informed consent for continuation of drug</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The subjects included in a LTE study should update consent each year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do you foresee other ethical considerations that should be considered in the document?
21. And finally, are there any remaining issues you would like to raise?

Thank you very much for your input!

We promise that next surveys will be much much sorter!
Dear Taskforce Member,

Please find below the second survey on the design and reporting of Trial Extension Studies (TES) as part of your task as a panel member.

In the first survey it was suggested that the length of an extension study varies and may depend on the length of the previous RCT; for example, an extension study following a short RCT would not necessarily be a long-term study. Accordingly, we have renamed ‘Long-term Extension Studies’ into ‘Trial Extension Studies’.

In this survey you will be provided with the results of the first survey which you can view before answering each question. Please read them carefully before answering the new question. There will also be the opportunity to include free text and comments to almost all the questions.

We would like to emphasise that there are no absolute wrong or right answers. The survey is reasonably detailed and is likely to take approximately 30 minutes to complete. Please make sure that you do not leave the survey half completed.

This survey is not anonymised - at the end of the survey, we will ask you for your email address in order to be able to send you additional information +/- feedback of your answers to this survey if indicated.

Maya Buch, Maarten Boers, Lucía Silva, and Loreto Carmona
Trial Extension Study (TES) Taskforce

1. PARTICIPANTS

In the first survey, we asked your opinion on involvement in the taskforce by representatives of industry, regulatory agencies (FDA, EMA) and contract research organisation (CRO). The responses were as follows:

A) Industry representative: 25% no involvement, 19% limited, 44% some, and 13% full.

B) FDA/EMA representative: 6% limited, 50% some, and 44% full.

C) CRO representative: 19% no involvement, 31% limited, and 50% some.

There is general agreement for some level of participation; therefore as a minimum, we will try to have a representative from the above mentioned bodies to provide comments on the final document; however, please rate your level of agreement for the following (0 = none to 10 = maximum)

<table>
<thead>
<tr>
<th>having an industry representative providing comments on one or more drafts</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>having an administration (FDA, EMA) representative directly participating in the taskforce</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>having a CRO representative providing comments on one or more drafts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Trial Extension Studies (TES) in Rheumatology

2. ITEMS TO BE INCLUDED IN THE FINAL DOCUMENT

Following your responses, the following items are to be included in the final document:

- A reference to the STROBE guidelines for reporting observational studies (mean agreement 9.3/10)

- A definition of TES studies (mean 9.7)

- A paragraph on the strengths and limitations of TES Studies (mean 9.1)

- A checklist with the minimal dataset a study should collect (mean 9.4)

- Recommendations on how to build a flow chart of the study (mean 8.4)

- Recommendations on how to analyse the data (mean 9.1)

- Recommendations on how to report the results (mean 9.2)

- Ethical comments (mean 7.9)

Do you agree with adding all the above items?

☐ Yes
☐ No

If no, please explain why
3. STUDY DESIGN DEFINITION

From the responses in the first round, no single definition for a TES proved to be satisfactory; please see the responses below:

“A TES may be a study that...

• ...follows patients beyond the pre-specified RCT period (mean 7.1)
• ...follows patients beyond the pre-specified RCT period, on condition that all patients entering the original trial are accounted for (mean 7.9)
• ...follows a placebo-controlled study (mean 5.0)
• ...follows an active-comparator study (mean 5.1)
• ...follows a placebo-controlled study in which patients start the experimental treatment either as a result of randomised allocation or crossover after placebo treatment; and only patients on experimental treatment are followed beyond the pre-specified trial period (mean 4.6)

We felt the best and most comprehensive definition was (mean 7.7):

“A TES may be a study that follows all patients beyond the pre-specified trial period whether the trial was:

A) a placebo-controlled RCT with the possibility to cross-over to open label drug;

B) OR a placebo-controlled RCT with the possibility to cross-over to usual care;

C) OR an active comparator trial “

Do you agree with the above definition?

☐ Yes

☐ No

If no, please explain why
4. STARTING POINT OF A TES

No clear definition of what constitutes the starting point of a TES emerged from the first round responses. We have added further options which we would like you to answer. Note that choosing a definition of the starting point does NOT imply data preceding that point should not be reported.

Which of the following options would you choose as the start point of a TES? (Choose only one)

- Start is on completion of the pre-specified RCT period (1st round 46% in favour)
- Start is at the point of cross-over or switch of therapy as allowed in the RCT (1st round 54% in favour)
- The start of the TES is the start of the extension trial (new item)
- The start of the TES depends on the research question (new item)

If you chose the last option, please specify:

5. LENGTH OF A TES

In the first round, most respondents supported not defining a minimum period of a TES (47%), and 40% voted for a period exceeding 3 years. Only 7% favoured up to 12 months and 25-36 months and no-one supported 13-24 months. We would like you to answer the following (which includes some additional items).

Which of the following options would you choose as the appropriate length of a TES? (Choose only one)

- More than 3 years (1st round 40% in favour)
- At least 5 years (new item)
- The minimum length of a TES need not be defined (1st round 47% in favour) as the length of a TES depends on the research question (new item)

If you chose the last option, please specify:
Trial Extension Studies (TES) in Rheumatology

**Population**

6. From responses in the first round, a single definition for the population to be included in a TES could not be decided. The responses were as follows:

“The population of a TES...

- Should include all patients initially included in the RCT (mean 9.1)
- Should include only those patients who achieved remission during the RCT (mean 2.0)
- Should include only those patients who achieved remission or a low disease-activity (LDA) state during the RCT (mean 2.8)
- May include patients with any level of disease activity on completion of the RCT (remission, LDA or moderate-high disease activity) (mean 7.4)
- Should not be defined upfront for all TES (suggested new item)

The best definition from feedback would seem to be:

“The population of a TES should not be defined in a guideline since it depends on the research question; but it should preferably include all patients initially included in the RCT (with the ability to separately report on patients with specific area of interest e.g. remission, LDA etc) “

**Do you agree with the above definition?**

- [ ] Yes
- [ ] No

If no, please explain why
7. Following the responses from the first round, the following items are to be included in the final document:

- Rare safety events may not be detected for reasons other than the length of observation (mean 7.8)

- Since TES are largely supported by pharmaceutical companies, the potentially limited access to data linkages (which are important for long-term observations) may further question the overall benefits of such studies (mean 7.1)

- Even if all patients in a RCT were entered into a TES, such a study is generalisable only to patients with similar disease characteristics; many trial populations do not reflect patients seen in routine care (mean 7.8)

- The definition of comparator groups in a TES may be difficult because of the absence of a clear null hypothesis (mean 7.4)

- Unwanted heterogeneity may result by wishing to accommodate countries where treatment options may be more limited (e.g. allowing patients with higher levels of disease activity to be recruited where otherwise inclusion is of patients in remission/LDA state only) (mean 7.6)

Do you agree with adding all of the above items?

- Yes
- No

If no, please explain why
8. One participant suggested the following in the first round:

It is desirable that TES are performed independently from companies (suggested new item)

How strongly do you agree with adding this item to the final document?

0  1  2  3  4  5  6  7  8  9  10

Space for comments

9. POTENTIAL SOURCES OF BIAS OR LACK OF GENERALISABILITY OF TES

From the responses in the first round, the following items are to be included in the final document:

“The inclusion of patients in a TES following completion of a RCT...

• ...usually requires a certain level of response (mean 7.9)

• ...may be influenced by the fact that the investigator is remunerated for each patient recruited (mean 7.4)

• ...may be influenced by geographical differences in practice / approach (leading to differences in number and nature of patients included) (mean 7.5)

• ...may be influenced by the stage of the disease of the patient (mean 7.0)

Do you agree with adding all the above items?

Yes
No

If no, please explain why
10. SAFETY

From the responses in the first round, the following items are to be included in the final document:

“A trial extension study...

• ...may identify new adverse effects that the original RCT was not able to detect due to greater cumulative exposure to the drug (mean 8.4)

• ...is not powered for rare adverse events and so should not be relied upon to detect safety signals (mean 7.6)

• ...may identify whether the incidence of known adverse effects show change with longer-term drug exposure (e.g. infection risk) (mean 7.5)

• ...may confirm whether the nature of known adverse effects identified from the RCT changes with longer-term exposure (e.g. infection risk) (mean 7.6)

• ...is inappropriate to detect rare safety events due to the inclusion of a selected population (responders with likely no previous adverse events) (mean 7.0)

Do you agree with adding all the above items?

☐ Yes
☐ No

If no, please explain why
11. EFFICACY

From the responses in the first round, the following items are to be included in the final document:

- The greater cumulative exposure of the active drug in a TES may identify additional information on the drug's efficacy (mean 6.9)

- A TES may allow evaluation of relapse including time to relapse (for example, patients entering the TES having achieved an acceptable state, such as LDA or remission, on the active drug can be followed to assess sustainability) (mean 7.8)

Do you agree with adding these items?

☐ Yes

☐ No

If no, please explain why
12. ADDITIONAL OUTPUTS

The responses from the first round regarding the additional outputs that a TES may provide were not conclusive. Indeed, there were more responses in the “disagree” or the “uncertain” area than “in favour” as shown below:

Besides efficacy and safety, and despite a clear scientific benefit, a TES may allow...

• ...economic evaluation of long-term treatment with the active drug (mean 6.3)
• ...better Health-related quality of life (QoL) analysis (mean 6.5)
• ...better evaluation of the risk-benefit ratio and therefore overall advantage of the drug (mean 6.9)
• ...information about compliance (suggested new item)

As a result, the introductory statement has been changed from “may allow” to “may also address”. Please answer now:

For the following statements on justifiable outcomes of a TES, please grade your level of agreement (0= none to 10= maximum)

Besides efficacy and safety, and despite a clear scientific benefit, a TES may also address...

- ...economic evaluation of long-term treatment with the active drug
- ...better Health-related quality of life (QoL) analysis
- ...better evaluation of the risk-benefit ratio and therefore overall advantage of the drug
- ...information about compliance (new item)
13. From the responses in the first round, the following minimal information a TES should collect will be included in the final document:

- The time of last observation (mean 9.5)
- The functional status at the time of inclusion in the TES (mean 8.8)
- The functional status at last observation (mean 8.3)
- The disease activity at the time of inclusion in the TES (mean 9.3)
- The disease activity at last observation (mean 9.4)
- The reason for exclusion from the TES if the patient discontinues the drug (mean 8.8)
- The reason for cessation of follow-up (mean 9.4)
- Specification of reasons for cessation of follow up other than adverse event or inefficacy, e.g. geographical or doctor related reasons (mean 8.7)
- The progress at each stage from RCT start to TES completion (mean 8.7)
- The duration of active treatment (mean 9.5)

Do you agree with adding all the above items?

- [ ] Yes
- [ ] No

If no, please explain why
14. A participant in the first round suggested the following items be added to the minimal information a TES should collect. Please state your level of agreement (0= none to 10= maximum):

- The co-medication at each stage from RCT start to TES completion
- The serious adverse events and any outcome related to safety at each stage from RCT start to TES completion

<table>
<thead>
<tr>
<th>Item</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>The co-medication at each stage from RCT start to TES completion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The serious adverse events and any outcome related to safety at each</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15. From the responses in the first round, the following item will be included in the final document:

"The minimum data requirements for TES following placebo- and active-controlled trials are the same." (93% in favor)

Do you agree with adding the above item?

- Yes
- No

If no, please explain why
16. From the responses in the first round, the following item is to be included in the final document:

“A TES that follows an active-comparator RCT should follow all randomised patients for the same period of time (not only patients on the experimental treatment)” (mean 7.9)

Do you agree with adding the above item?

☐ Yes
☐ No

If no, please explain why
17. From the responses in the first round, with most of the items strongly supported, the following statements regarding the minimal information that a TES should collect are to be included in the final document:

- The null hypothesis should be stated at the start (mean 7.9)
- Multiple comparisons should be taken into account when determining the level of statistical significance (mean 8.1)
- The report should include details on how data for sustained effect was analysed (mean 9.4)
- The plan for subjects that drop out of a TES should be specified (mean 8.9)
- In a TES, the planned analysis of data to evaluate for sustained effect should be non-inferiority in nature (mean 7.6)
- The analysis of the data from a TES in rheumatoid arthritis (RA) should include the area under the curve of absolute disease activity (i.e. not response/change) preferentially expressed as a score (DAS, SDAI, etc.) (mean 7.3)
- The analysis should include survival / retention rates (mean 8.9)
- A TES should preferably include hard endpoints (e.g. death, work disability, joint replacement surgery, hospital admission) from linkages with other data sources (mean 8.6)
- The analysis of the data from a TES should take into account the dropouts (mean 9.3)

Do you agree with adding all the above items?

☐ Yes
☐ No

If no, please explain why
18. Two additional items were not clearly supported, although by the comments made and number of non-responses, the statements may not have been clear to the participants.

- The null hypothesis must be related to the results achieved in the original RCT. (mean 6.3)

- The analysis should include a pooled analysis from the original trial groups. (mean 6.4)

The items have therefore been re-phrased as below. Please rate your level of agreement with the following items (0 = none to 10 = maximum):

<table>
<thead>
<tr>
<th>Item</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>The null hypothesis should take account of the results of the original RCT. Depending on the research question, the results of a RCT should be accommodated in the TES.</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>The report should comment on cumulative outcome analysis (beneficial and adverse events) maintaining the original trial groups i.e. from RCT start, not TES start to avoid reporting of only the sub-selected patient group that proceeds onto the TES.</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
</tbody>
</table>
Patients randomised to either group of the RCT may leave the trial early. Patients on the active arm may continue receiving the experimental drug until the end of the RCT or discontinue it before ending. Patients on the placebo/comparator group may continue on placebo/comparator until the end of the RCT or start on the experimental drug before the end of the RCT. At the start of the extension period two subgroups may therefore be receiving the experimental drug; in addition, those continuing in the placebo/comparator arm may also commence experimental drug if they enter the TES phase. During the TES, patients may continue receiving the experimental drug or discontinue it before completion.
19. From the responses in the first round, the following item is to be included in the final document:

“All TES reports must include a flowchart on the progress of patients included” (mean 9.9)

The previous flowchart (agreement mean 9.3) has been improved following participants’ comments: specifically the number of withdrawals during the TES has been added and some of the groups re-labelled.

Please rate your level of agreement with the revised flowchart (0 = none to 10 = maximum)
20. From the responses in the first round, it has been decided that the taskforce should develop and detail minimal standards that should be included when reporting a TES by means of a checklist (mean 9.2)

The following items are to be included in the final document:

• “The report of a TES should be consistent with and consolidate existing established guidelines including CONSORT and STROBE” (mean 9.4)

• The report of a TES should be consistent with the ACR/EULAR recommendations on the reporting of clinical trials in RA (Aletaha D, et al 2008) (mean 8.9)

• A report of a TES should include a flow diagram detailing numbers at each relevant time-point (mean 9.5)

• For those patients entering the TES having achieved LDA or remission during the RCT, the sustainability of such disease states should be evaluated and made available (mean 8.6)

• For those subjects that enter a TES not having achieved remission/acceptable disease activity state following the RCT, the number that achieve this during the TES should be reported – to determine whether longer drug exposure has the potential to improve disease state of such subjects further (mean 8.3)

• The drop-out rates from each arm during the original RCT and the cross-over groups should be available (mean 9.3)

• All drop-outs should be detailed (mean 9.1)

Do you agree with adding all the above items?

☐ Yes
☐ No

If no, please explain why
21. A participant in the first round suggested the following item be added:

- TES do not need to provide the same information for all randomised patients in terms of quantity and quality, irrespective of the nature of prior RCT, but should provide a minimum dataset (still to be defined) for the patients not followed in detail. Additional information, depending on the research question, is desirable but not mandatory.

How strongly do you agree with adding this item (0 = disagree to 10 = fully agree)?
22. From the responses in the first round, we have concluded that the taskforce document should comment on the frequency (mean 7.6) and nature (mean 8.0) of the reports of TES.

Regarding the nature of reports, the following statements are to be included in the final document:

- The results of efficacy and safety of a TES should be reported together (mean 8.8)

- The credibility of split reporting (e.g. one abstract on efficacy, one on safety, one on QoL etc) – is questionable and should be discouraged by abstract selection committees and journal editors (mean 8.7)

Do you agree with adding all the above items?

- Yes
- No

If no, please explain why

23. A participant in the first round suggested the following item be added:

“The results of the groups of patients in different disease states should be reported separately”

How strongly do you agree with adding this item (0 = disagree to 10 = fully agree)?
24. Regarding the frequency of reports, the responses were not clear or not supportive of any specific alternative:

- The results of a LTE study should be reported every year (mean 4.2)
- The results of a LTE study should be reported every two years (mean 4.9)
- The results of a LTE study should be reported every three years (mean 5.4)
- The results of a LTE study should be reported annually to a maximum of 5 years (mean 5.4)

The following statement has therefore been decided to be included in the final document:

“The reporting frequency should not be specified for all TES since it depends on the research question. However, the protocol of each TES should specify the frequency of reports to be written and a reason for it (purpose, outcomes, length of RCT, etc).”
25. From the responses in the first round, the following items are to be included in the final document:

- All of the subjects undergoing a RCT should be informed of the importance of long-term surveillance and be given the opportunity of entering in the long-term follow-up (mean 9.4)

- The subjects included in a TES should sign a new informed consent form (different from the one for the RCT) for continuation of data collection (mean 7.6)

And the following item is to be deleted:

- The subjects included in a TES should sign a new informed consent for continuation of drug (mean 5.9)

Do you agree with the above changes?

☐ Yes
☐ No

If no, please explain why

26. In addition, a participant in the first round suggested the following item be added:

- There is no need to update the consent of patients included in a TES annually, particularly since each additional consent runs the risk of additional drop-out

How strongly do you agree with adding this item (0 = disagree to 10 = fully agree)?
27. We wish to be able to send responses to participants directly as may be indicated; therefore, please include your e-mail address below: