ISPOR
The professional society for health economics and outcomes research

Improving healthcare decisions

505 LAWRENCE SQUARE BLVD SOUTH
LAWRENCEVILLE, NJ 08648

P +1-609-586-4981
F +1-609-586-4982

info@ispor.org

30 September 2023

EMA/CHMP/564424/2021

Dear European Medicines Agency (EMA):

ISPOR – The Professional Society for Health Economics and Outcomes Research - is pleased to respond on behalf of its membership to your consultation entitled "Reflection paper on establishing efficacy based on single arm trials submitted as pivotal evidence in a marketing authorisation."

ISPOR is a scientific and educational society with many of its members engaged in evaluating health technologies, including pharmaceuticals, medical devices, and other interventions. We have a large membership living and working in 110 countries globally, across a range of disciplines, including health economics, epidemiology, public health, pharmaceutical administration, psychology, statistics, medicine, and more, from a variety of stakeholder perspectives, such as the life sciences industry, academia, research organizations, payers, patient groups, government, and health technology assessment bodies. The research and educational offerings presented at our conferences and in our journals are relevant to many of the issues and questions raised in this request for information.

The response to this consultation was led by the Policy Outlook Committee of our most senior advisory body, the Health Science Policy Council. To engage our membership, we consulted with interested members of our Institutional Council (ie, industry and consulting), as well as our Real-World Evidence, Rare Disease, Statistical Methods in Health Economics and Outcomes Research (HEOR), Oncology, Precision Medicine and Advanced Therapeutics, Patient-Centered, and Medical Devices and Diagnostics Special Interest Groups. The attached document provides both summary and line-by-line responses based on their comments. We hope they prove useful.

ISPOR would be happy to answer any questions about our response, and to participate in any follow-up consultations on the relevant program items mentioned within the report.

Sincerely,

Robert Abbott
CEO & Executive Director
ISPOR

30 September 2023

Submission of comments on ' Reflection paper on establishing efficacy based on single arm trials submitted as pivotal evidence in a marketing authorisation' (EMA/CHMP/564424/2021)

Comments from:

Name of organisation or individual
ISPOR – The Professional Society for Health Economics and Outcomes Research We would like to acknowledge ISPOR members Grammati Sarri and Allicia Girvan for their assistance in assembling these comments, as well as ISPOR staff Richard Willke, Laura Pizzi, and Kelly Lenahan.

Please note that these comments and the identity of the sender will be published unless a specific justified objection is received.

When completed, this form should be sent to the European Medicines Agency electronically, in Word format (not PDF).

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

ISPOR
The professional society for health economics and outcomes research

Improving healthcare decisions

505 LAWRENCE SQUARE BLVD SOUTH
LAWRENCEVILLE, NJ 08648

P +1-609-586-4981
F +1-609-586-4982

info@ispor.org
www.ispor.org

General comments

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	The reflection paper on Single-Arm Trials (SATs) is a foundational guide that offers a comprehensive overview of the regulatory framework surrounding SATs. However, the document could be significantly enriched by incorporating actionable insights, real-world examples, and economic considerations. These elements are particularly relevant to Health Economics and Outcomes Research (HEOR), a field that increasingly relies on SATs for evidence generation, especially in the context of orphan drugs and personalized medicine. Although this paper indicates (ll. 352-353) that external data are beyond its scope, the use of external control arms has received considerable attention of late (eg, a recent draft guidance from the FDA). Indeed, some of the considerations noted later in the document cannot be interpreted without the reference to the origin and quality of control (external) arm including the type of statistical analysis. The ICH E10 Choice of control group in clinical trials (CPMP/ICH/364/96) from 2001 is of course still quite relevant, as are some other guidelines listed in Section 2, but a reflection paper that pulls together recent thinking about external control arms (including methodological approaches that such as an external arm that is informed by real-world data or employs target trial emulation, alluded to but not explicitly discussed in section 4.3) would seem to be a critical complement to the present paper. In several instances the draft paper makes subjective-type references that will not allow consistency in "translating" this paper, not only among readers but	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	also during the assessment process (examples, line 42 “fraught with shortcomings”, line 148 "leaves no doubt", line 153 "truly impossible, line 214 "negligible extent", line 295 "impossible to validate"). Taken together, they create what comes across as a generally skeptical tone about the validity of SATs in much of this paper. It may result in a general hesitancy, especially from industry & payers, to be the first use case of a new approach as there are many unknowns. However, EMA has approved several products based on SATs which demonstrates the value of this study design for indications with high unmet need/rare and very rare indications when randomized controlled trials (RCTs) are not feasible or unethical. Please see a review on SAT-based approvals for oncology drugs (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10163162/). Figure 1 in this publication shows the trend of increasing SAT acceptance over time by EMA. These cases could provide useful examples for the validity of SATs in general as well as specific approaches to concerns described in this paper. We would suggest a more explicit discussion of diagnostic technologies and medical devices. SATs are commonly used to estimate efficacy of devices in particular. If the authors are looking to broaden the impact of this paper to these areas, we suggest they include some background from medical devices and diagnostics experience related to SATs and marketing authorizations. There are reasons there exist more SATs for devices (eg, ethical considerations and difficulty in providing a safe “sham” procedure, etc.). To that point, many of the points the EMA makes (eg, trial population and endpoint selection, prior trial and analysis specification, etc) are mature and common practice among device trials.	

Specific comments on text

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
33		**Comment:** Given the observation in ll. 27-28 that a "relevant proportion of marketing authorization applications … stems from single-arm trials", why is "defining general conditions under which SATs may be considered acceptable" outside the scope of this reflection paper? **Proposed change (if any):** Please add a section that defines the general conditions under which SATs may be considered acceptable as pivotal evidence for marketing authorization; this seems very much in scope for this paper.	
36		**Comment:** It is unlikely that SATs can provide "clear" pivotal evidence of if "clarity" is based on having a randomized control arm. Please consider referring to "acceptable" or "sufficient" instead of clear, which is later used in the paper. That said, acceptability must be assessed with reference to the control (external) data. One could argue that this is also the case for RCTs when there are serious issues with the control arm (lack of representation of clinical care, high level of missing data, etc) and no comparative effectiveness can be established. **Proposed change (if any):** Please consider referring to "acceptable" or "sufficient" instead of clear, which is later used in the paper.	
43		**Comment:** While this reflection paper explicitly related to "establishing efficacy", a SAT will also yield important safety-related evidence, as indicated here. Both the complementarity and tradeoffs (eg, in inclusion/exclusion criteria) of generating both efficacy and safety evidence in an SAT may be worth some discussion.	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		Proposed change (if any): Consider a broader discussion regarding the interplay of considerations when including both efficacy and safety endpoints.	
58-66		**Comment:** The brief description of SATs in this section misses the opportunity to reference key considerations related to the specific features of this study design. Although it is acknowledged that references to external (control) data is beyond the scope of this reflection paper, this section cannot be completed without noting the different sources of the control arm that can define the type of SATs (for example use of historical or concurrent control, externally collected, data). This section can be improved by specifically referencing to the "The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology (Revision 11)" (page 91) that will guide readers to gain more information about defining different types of non-randomized trials. The paper briefly mentions that (line 144) without extrapolating how different methodological considerations in planning prospectively or selecting the "right" comparator (external) data may allow for reliable causal interpretation of treatment effect. **Proposed change:** see suggestions above.	
68		**Comment:** Must SAT's lack a "concurrent control arm"? Sometimes an external control arm run concurrently with the SAT is feasible in a registry or electronic health record (EHR). **Proposed change:** Consider replacing "concurrent" with "randomized."	
105-201		**Comment:** This section is set up to articulate the definitions around the main points for the design, planning, conduct, analysis, and interpretation of SATs without	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		reference to the "key" differentiator of this study compared to the RCTs; which is the absence of the control (or alternative treatments) arm. Although this section provides details on key methodological topics that may impact uncertainty in treatment effect estimates, several of these topics are not strictly related to the study design of SATs but generally applying to any study design (including RCTs). The draft paper refers to the ICH E9(R1) guidance and the concept of estimands and makes clear connection between the appropriateness of a SAT and whether it can address the targeted estimand of interest (line 130). However, one element for defining these estimands is the consideration of comparators. The central question for these concepts is how the outcome of the treatment tested in the SAT compares to what would have happened to the same subjects in an alternative scenario (i.e., had they not received the treatment, or had they received a different treatment). We again find these topics lacking direct applicability for establishing efficacy in SATs without bringing the considerations around the use of external (control) arms or without adding specific references from other guidance documents/ key papers that can help readers to follow appropriate methodology in this area. **Proposed change**: In general, please consider expanding the paper, or producing a separate paper, to discuss many of these same points when using an external control arm.	
203-207		**Comment:** To follow up on our general comment about devoting more attention to medical devices and diagnostics, there are multiple considerations relevant to the design of the evidence generation to demonstrate effectiveness. Among others: • Type of device: Therapeutic: surgical, orthopedic, neurology,	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		esthetic; Diagnostic: companion diagnostic, invitro diagnostic, imaging diagnostic, genetic, etc.; Therapies: Digital diagnostics, Technologies, and diagnostics. • Unique Features: Device complexity, availability (regional) user effect (training and user skills), learning curve, and additional human factors among others to consider. Depending on the previous considerations, sponsors of clinical trials need to assess additional factors for designing a clinical study. Among them: • Study Objectives • Subject Selection • Stratification of the subject selection • Site selection (and availability of comparison) • Study Design • Assessment of the bias and treatment effect Due to their complexity devices and diagnostics require an additional level of considerations for the design of the adequate type of study to demonstrate its effectiveness (either therapeutic or diagnostic) much broader than pharmaceutical world. **Proposed change:** Please consider including points and examples that are specifically relevant to medical devices and diagnostics.	
203		**Comment:** Consider replacing "suitable"' with "desired", since RCTs are not always feasible/suitable. In general, RCTs are the most desired method to provide reliable estimates of clinical efficacy. **Proposed change:** see comment.	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
204		**Comment:** "In certain situations, evidence from SATs may be considered acceptable." A summary of those situations would be very helpful such as when RCTs may be challenging to conduct due to ethical considerations, in diseases with high unmet need with difficulties in patient recruitment, or when early-phase trial data for conditions with no established standard of care reveal large efficacy benefits. **Proposed change:** see comment.	
214-216		**Comment:** Regarding "it is required that the primary endpoint is such that observations of the desired outcome would occur only to a negligible extent …" - doesn't this depend partly on whether a valid, well-matched external control arm (ECA) exists and careful quantitative bias work is done? On another note, it could be useful to clarify that the opposite would be just as good, i.e., that it (almost) always occurs without treatment. **Proposed change:** see comment.	
252-265		**Comment:** In the literature a "multi-domain responder index" has been proposed for rare disease programs. By incorporating different domains, the causality assessment for such an index may be more questionable than a usual continuous endpoint, and hence (lacking a strong ECA) the use of such an index should be carefully scrutinized in a SAT. **Proposed change (if any):** Add a paragraph to address "multidomain responder index" as a special case of continuous endpoints.	
266-277		**Comment:** There is an additional type of	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		binary endpoint, which is a responder endpoint where a responder status can be achieved by achieving any of a few different components, eg, a collection of different development milestones. By having multiple components, it may be particularly difficult (lacking a strong external control arm [ECA]) to assess if a change could have occurred without intervention. Thus, this type of endpoint could also be discouraged in a SAT. **Proposed change (if any):** Add a paragraph to address such binary composite endpoints as a special case of binary endpoints.	
202-282		**Comment:** The section could also be enriched by discussing quality of life endpoints, and perhaps economic endpoints for subsequent health technology assessment (HTA) use. It would also be good to see more patient-reported/patient-centered outcomes captured within trials overall. SATs could include those with known minimal clinically important differences (MCIDs) for a population where a control does not exist. Quality-Adjusted Life Years (QALYs) for use in cost-effectiveness analyses could be cited as a relevant economic endpoint in SATs. **Proposed change:** see comment.	
292-297		**Comment:** These considerations are not strictly bonded to the use of SATs and can apply to any study design (even RCTs) for rare and very rare conditions with high heterogeneity lacking clarity regarding the natural disease history. We agree that these considerations may be stronger for SATs lacking a control group but are not unique for this study design. **Proposed change:** please consider acknowledging the generality of this issue in	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		this paragraph.	
335-337		**Comment:** External information to compile the "control arm" and derive a comparative treatment effect for SATs is crucial. However, this section is very brief and lacks information to capture the breadth of topics around use of external information and its impact in SATs; for example, different sources to select external information (concurrent vs non-concurrent external controls, different types of non-concurrent data such as retrospectively collected natural history, published data, previous clinical study, baseline-controlled study). Each ECA data source has unique methodological challenges that need to be considered in the SAT planning and analysis. Some of these methodological topics are ECA specific; for example, different types of biases are applicable in one study design but not in others. Furthermore, no reference is made in this section regarding the selection process for external control data; for example, using a systematic approach of identifying the availability of relevant external data and appraising the fitness-for use of different data sources including the patient comparability profile with those included in SATs. This process is far from straightforward with unique challenges and trade-offs between selecting the most comparable and appropriate external control arm (especially considering that most often SATs are tested in small populations with lacking available established clinical practice and clinical expertise) therefore more importance should be placed on how to transparently documenting these decisions through planning and analysis of SATs. We strongly believe that if this paper does not aim to comprehensively cover the topics around external control arms, reference to other already published guidance in this section	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		(for example ICH E10) is crucial to give necessary information to the readers. **Proposed change:** see comment	
358-488		**Comment:** The section is comprehensive but should include additional practice guidance for HEOR analysts, including software like SAS and R that are commonly used for statistical analyses in SATs. **Proposed change:** see comment	
361-362		**Comment:** We suggest including guidance for situations when a lag time exists between publications and new studies beginning prior to public knowledge of previous study results. **Proposed change:** see comment	
450		**Comment:** We suggest avoiding use of the term "unambiguously". **Proposed change:** see comment	
483-488		**Comment:** We suggest that additional clarification be provided regarding sample size in SATs and testing in general (eg, predefined threshold or precision in terms of confidence intervals). **Proposed change:** see comment	
469-475		**Comment:** Although we agree with the statements in the section, the following proposed changes are recommended. **Proposed change:** Replace "in some settings" with "in most settings", and "might" with "should". This is analogous to specification of a fixed margin, where the effect of the active comparator should be based on a confidence limit rather than a point estimate.	
489-497		**Comment:** This section could be more informative by providing real-world	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		examples. **Proposed change:** For instance, the potential for "selection bias" in SATs could be discussed, along with mitigation strategies like propensity score matching.	
457-459		**Comment:** Given the discussion in the "Choice of endpoints" section, this sentence would make more sense if "negligible" were replaced by "non-negligible." **Proposed change:** Please clarify the point above.	
484		**Comment:** A clearer definition of adequate sample size is recommended. Additionally, guidance is recommended in circumstances when the required sample size is not feasible (eg, rare diseases). **Proposed change:** please specify statistical references.	
498-499		**Comment:** This is a very useful table; the fact that bias reduction measures exist should be acknowledged more often in previous sections. More discussion of quantitative bias approaches would be useful (eg, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4984770/). **Proposed change:** see comment	

Please add more rows if needed.