**EFSPI Meeting with MHRA Statisticians Minutes**

**8th November 2023**

**Attendees**

MHRA: David Brown, Julia Saperia, Khadija Rantell, Maria Peppa.

EFSPI: Yolanda Barbachano, Anna Berglind, Michael Cartwright, Anne Danniau, Lars Endahl, Dan Evans, Chrissie Fletcher, Christoph Gerlinger, Tiina Hakonen, Jürgen Hummel, Rima Izem, Mette Krog Josiassen, Fredrik Öhrn, Giuseppe Palermo, Alessandro Previtali, Karen Smith, Florian Voss, Anja Von Heydebreck.

|  |  |
| --- | --- |
| **Topic** | ***Questions* and Notes** |
| **Opening** | The meeting opened with the introduction (name, company and location) of the attendees from both MHRA and EFSPI |
| **Follow up from last meeting** | *Is there any follow up on any items discussed at the last meeting (23 November 2022) not covered elsewhere?*  EFSPI highlighted that a new statistical methods leaders group has been established within EFSPI and will wish to engage with regulators in future. |
| **What is on your agenda?** | *Can you give us an update on current statistical issues and potential regulatory statistical concerns of the future?*  *What regulatory guidelines are planned by MHRA? Where do we see diversion in regulatory thinking from EMA (eg single arm reflection paper by EMA)?*  There is no guidance planned (that the MHRA statisticians are aware of) that will diverge from EMA. The MHRA team highlighted that the UK has now been admitted as an independent country to ICH (previously provided their input via the EMA).  The MHRA are continuing to focus on Real World Data topics and are currently working on external control arms, and a draft guideline is expected to be released for consultation next year. They are also involved in an MHRA led project on synthetic control arms (<https://www.gov.uk/government/publications/projects-selected-for-the-regulators-pioneer-fund/projects-selected-for-the-regulators-pioneer-fund-2022> ).  MHRA statisticians also participate at various European innovative medicine initiative and innovative health initiative (IMI/IHI) (https://www.ihi.europa.eu/) research projects such as SISAQOL-IMI (<https://www.sisaqol-imi.org/> ). |
| **Intercurrent events of treatment switching and its handling for long-term outcomes** | *The ICH-E9 addendum highlighted the importance of specifying the estimands and ensuring that this target guides design and analytical choices. Eliciting intercurrent events, assessing how incorporate them in the question of interest, and planning how to address them in the design and analysis are important tasks in this specification. Change of exposure status during follow-up (e.g., stopping treatment, taking rescue therapy, switching, adding a concomitant medication) are potential intercurrent events, especially when these events precede a longer time of assessment of the variable of interest. An estimand defined with a treatment policy strategy to handle these intercurrent events may be easy to specify but hard to interpret.*    *Questions:*   * *What type of arguments, whether statistical or clinical, is the MRHA looking for in the protocol or the SAP for justifying estimand strategies (different from treatment policy) for handling intercurrent events such as switching?*   The MHRA does not have a specific set of acceptable arguments for justifying the estimand strategy. MHRA team emphasised the importance of clarity; being clear on what we are trying to estimate and explaining why what is being done is appropriate way to achieve that. It’s important that the estimator and the target estimand are aligned. It is acceptable to have a different approach for each intercurrent event. The MHRA expect clinical input and consideration of the disease setting when selecting a particular strategy for handling an intercurrent event.  There has been a noticeable improvement in implementation of the estimands framework; which is evident both during Scientific Advice consultations and trial protocols review, although some protocols are still quite sparse on the details of the clinical questions of interest and description of estimands’ attributes. There remains some lack of clarity when working backwards from the handling of intercurrent events and the exact question that these are attempting to address. There are increasingly many more regulatory guidances that clearly outline their preferred estimands and the strategies for handling certain current events.   * *Are there special circumstances where such arguments are discouraged, and one estimand strategy is clearly preferred? Or are there special circumstances in which these arguments are welcome and encouraged (e.g., non-inferiority testing)?*   The MHRA team indicated that there were no preferred estimand strategies and that the choice of estimand depends upon specific study objective. The overriding concern is to ensure that the estimation of treatment effect is conservative. They expect explanations of why a certain approach was used and what sensitivity analyses may be useful. It is important to consider the frequency and type of intercurrent events. Details of various differential patterns of discontinuation should be provided along with an assessment of their impact on the estimation of treatment effects.   * *Are there estimand strategies or estimators beyond the ones that were outlined in the ICH-E9 that could be considered? (e.g., use of multi-state models methods and handling of competing risk)?*   The MHRA’s short answer was that they haven’t seen many examples of methods beyond those outlined in ICH E9 (R1). The MHRA are open to new methods, as long as the justifications for their use are acceptable i.e. how are these methods aligned with the target estimand. The topic of robustness was also raised, and the MHRA team stressed that the initial focus should be on the robustness of what is being estimated for the primary estimand rather than the robustness between estimands and the clinical questions of interest. It is also reassuring to see whether the data collected are robust to many of the methods/assumptions that were originally planned for e.g., if very few, if any, patients actually experienced certain types of planned intercurrent events then which approach was used for handling that event becomes unimportant.  Another important topic is unplanned events impacting the conduct of a trial i.e. unforeseen trial disruptions. It is important to build in flexibility into the trial by design. The MHRA is a member of the Estimands Implementation Working Group (EIWG). The EIWG is trying to address a number of topics relating to the implementation of the estimands framework.  EFSPI raised the additional question of the acceptability of the consideration of Operating Characteristics when simply imputing extreme values significantly inflates variability.  The MHRA team answered that it would be good to demonstrate the sensitivity of this assumption to justify an alternate imputed value for example via the use of tipping point analyses which impute more extreme values. |
| **Subgroup analyses** | *Subgroup analyses of randomised clinical trials have received a lot of attention among academia, prescribers, payers, regulators and industry. The underlying rationale for requesting or conducting subgroup analyses rest on two fundamental assumptions:*   * *Treatment effect is heterogenous in the trial population* * *Within the trial population there are subgroups for whom the treatment effect is considerably more homogenous*   *Among payers the interest is often whether subgroups with particular good effect can be identified through intrinsic or extrinsic factors assessed in the trial population that will justify reimbursement only for particular subgroups. Among regulators the interest is often “to check that the estimated overall effect is broadly applicable to relevant subgroups” (cf. CHMP Guideline on the investigation of subgroups in confirmatory trials)*  *The statistical issues of assessing subgroup homogeneity/heterogeneity are well-known and include:*   * *Lack of a relevant criterion to assess subgroup heterogeneity* * *Lack of power in many if not most trials to detect subgroups with a differential treatment effect* * *Risk of false positive findings* * *Even if true differential effects are identified, they may be too small to matter in clinical practice - as assessed by e.g. Positive Predictive Values and Negative Predictive Values (see e.g.* [*https://academic.oup.com/jnci/article/107/8/djv153/950700*](https://academic.oup.com/jnci/article/107/8/djv153/950700) *for time-to-event example)*   *Question to MHRA:*   * *Are there specific situations and/or therapy areas where, in your experience, the two underlying assumptions for doing subgroup analysis are more or less plausible?*   The MHRA team used as an example to illustrate their thoughts the commonly encountered oncology genetic biomarker situation with +ve/-ve biomarker subgroups and the assumption that the drug will likely have its largest effect in the +ve subgroup. In this situation, they often find that the analyses presented by the Sponsor will omit the -ve subgroup and simply present the overall effects along with the +ve subgroup. Although it’s important to see the results in the -ve subgroup, statistical significance will unlikely be required. It’s important to remember that the key regulatory concern is to assess whether a positive benefit/risk is maintained across subgroups. There is less interest in interaction tests and the general direction of consistency is their main focus. They also look at other endpoints, rationale for the drug working in this subgroup, expected or surprise findings, etc. and all of this combined information is taken into account.  The EFSPI team raised the additional question of whether separate trials for the different populations might be of more interest in this oncology situation? The MHRA team felt that separate trials would likely be an unnecessary complication in trial conduct for this situation.   * *Among criteria often suggested for claiming differential treatment effects are significant p-value for the heterogeneity test, that the point estimate for a subgroup is outside 95%CI of effect estimate in the entire population or that the heterogeneity is clinically relevant (cf. ICH-E17). Which, if any, of these criteria are reasonable to use in regulatory decision making?* * *How does the lack of power to detect differential subgroup effects affect regulatory decision making?* * *Are there examples where subgroup heterogeneity has affected regulatory labelling/decision making at MHRA, and if so what analyses or evidence were used in the decision making?*   The MHRA team began their answer to these questions with the assumption that the context of these subgroups is, that there are standard subgroups that are often assessed to check consistency of results. In this context, they highlighted that the assessment of the consistency of results across the subgroups isn’t a key area of focus and only if extreme reversal of an effect is observed then some additional investigation may be warranted. They are cautious not to over interpret any variation in effects unless there are other signals or a clear rationale. It is unlikely that this would be of interest and additional statistical approaches would not be applied. MHRA don’t look to exclude subgroups but if this situation arises it is not just a statistical decision and in this situation a replicate phase 3 trial could be valuable (though whether two trials of sample size n are more informative on this point than a single trial with sample size 2n is unclear). If a replicate phase 3 trial is not available, they can look at previous data, e.g. phase 2, pre -clinical and see if there are any hints of this being a true effect. The MHRA do not find interaction tests to be an informative approach as they overreact in large trials and miss things in small trials. Their general approach is to assume chance finding unless all data points otherwise.  The MHRA sometimes see regional differences and these may warrant further investigation, but don’t generally react if EU/UK subgroup doesn’t look as good as US for example. Sometimes they can learn from it: e.g. if compliance is worse in UK, is it because of differences in the administration approach and might a co-administered food/flavour be different? Does the company need to change anything?   * *Should Bayesian shrinkage be default in all subgroup analyses where no prior mechanistic knowledge would suggest a differential effect?*   The MHRA expressed their preference to see the standard results initially, but if the Bayesian shrinkage argument contributes to interpretation, then this would be acceptable.   * *Is the ability to utilise differential treatment effects in clinical practice part of the regulatory decision making when it comes to assessment of subgroup effect estimates, and if so what type of calculations would support that e.g. a restriction in the indication excluding a certain subgroup would lead to better treatment options in the target* *population?*   The MHRA team indicated that almost always you would want to see both results in the different populations.  EFSPI raised an additional question about the use of the multiplicity argument as a valuable argument to explain isolated extreme results. The MHRA answered that that this may be an appropriate approach, but not if the subgroups have been pre-planned due to clinical rationale relating to possible different effects. |
| **Considerations about acceptability of novel endpoints for regulatory decision making** | *Development of new drugs can become extremely challenging in certain situations, e.g.*   * *for diseases which are progressing slowly (e.g. pre-symptomatic Alzheimer Disease (AD), Multiple Sclerosis (MS)), requiring a very long trial duration to show an effect on traditional endpoints, often resulting in suboptimal dose-finding, with a consequent increase in the risk of failing Phase-3* * *for diseases with endpoints which are not sensitive enough, even more if the disease is rare*   *In those cases, there is a clear need to develop novel and more sensitive endpoints to shorten the entire development process. With the ultimate goal of enabling earlier decision making, for patients to have a faster access to effective drugs.*  *Those novel endpoints can be of different nature, for example:*   * *new clinical outcome assessment (COA), like clinician-reported outcome (ClinRO), performance outcome (PerfO) or even COA captured by a digital and passive wearable device* * *biomarkers-based endpoints*   *and they could also be defined as a combination of multiple domains/components (composite endpoint).*  *A positive treatment effect on the novel endpoint could led directly to a full market authorization (e.g if the endpoint is qualified as primary endpoint, in a specific context of use, or as a surrogate), or constitute a meaningful intermediate endpoint, reasonably likely to predict the final endpoint, and thus potentially lead to an accelerated approval.*  *The drug-development field is quite active in the effort of developing new endpoints, just to mention a few examples:*   * *The Partnership and Alliance in Acute myeloid leukemia Clinical Treatment (MPAACT), an industry-led research alliance founded in 2018, is trying to establish the minimal residual disease (MRD) as a surrogate endpoint in Acute Myeloid Leukemia (AML)* * *The Alzheimer's Prevention Initiative, a public-private partnership intended to accelerate the evaluation of AD prevention therapies, is evaluating the API Preclinical Composite Cognitive Battery (APCC), a composite endpoint, as a primary endpoint for individuals at risk for clinical onset of Alzheimer's disease* * *The 95th centile of the stride velocity (SV95C), a COA captured by using a digital and passive wearable device, has been very recently qualified, by the Committee for Medicinal Products for Human Use (CHMP), to be used as a primary endpoint in confirmatory superiority studies, for ambulant Duchene Muscular Dystrophy (DMD) patients 4 years of age and above*   ***Some statistical issues related to this topic are:***   * *How to establish surrogacy? How to establish an ‘intermediate’ endpoint?* * *Acceptability of composite endpoints*   ***Questions to MHRA***   * *What’s the status of the Innovative Licensing and Access Pathway (ILAP) initiative and how it may be related to this topic?*   The MHRA highlighted that the idea of ILAP is that the regulator and HTA is involved in ongoing discussions and updates during a products development. Scientific Advice sessions are the forums where these topics are usually discussed. If a non product-specific view is required, this would have to occur outside of the ILAP process.  Nothing has come through ILAP yet in terms of new endpoints to qualify.   * *Do you have recent examples of novel endpoints, which were key in the approval process?*   The MHRA haven’t encountered any recent examples.   * *What is your view on the potential use of biomarkers to accelerate clinical drug development, for a faster access to drugs from patients?*   The MHRA team acknowledged that this is of great interest as we’re all interested in getting medicines to patients faster. However, they highlighted that the level of evidence required to validate a surrogate endpoint might be considerable. They highlighted that the use of the Early Access to Medicines Scheme might possibly be an appropriate route to discuss the use of biomarkers as proof of efficacy for early access on a case-by-case basis.   * *Is there a minimal amount of evidence required to establish surrogacy (e.g. minimum number of studies in a meta-analysis)? Which statistical approaches do you regard as the ‘key evidence’ for establishing surrogacy?*   The MHRA team answered that there were no specific requirements for the overall validation package, but it’s important to highlight that correlation within a single trial is unlikely to be sufficient and additional data sources will be required outside of the trial as correlation of treatment effects across trials are necessary.   * *Do you have examples possibly across indications, beyond the classical PFS -> OS, where novel endpoints were used as intermediate endpoints to gain early access to drug?*   The MHRA team answered no, not even in EAMS.   * *What do you see as the main challenges with developing composite endpoints?*   MHRA commented that they didn’t have a lot of experience with this topic. The choice of components to be included in the composite endpoint must be carefully chosen to reflect clinical benefit, particularly in relation to their relative importance and direction of effect.  The MHRA would likely have general concerns about the potential for an individual component to drive the overall effect in a composite. |

Two additional topics were raised by the MHRA and they were interested in EFSPI experience: the use of AI in designing clinical trials and the adoption of Bayesian approaches.

EFSPI felt that the AI tools available are still at a very early stage, but likely to see wider use in the very near future and this will obviously bring concerns with validation/interpretation etc and likely require the involvement of a wider group of individuals with specific skill sets outside of statistics. There may be applications in analysis of large data e.g. imaging to develop surrogates, but currently it feels like we are mainly using AI approaches to confirm findings which we are already aware of.

EFSPI summarised that Bayesian approaches are used extensively during trial design and often applied in early-stage decision making. Their use in paediatric extrapolation analyses is also another example of Bayesian approaches. The use of Bayesian borrowing in confirmatory trials is also proposed at the design stage and during Scientific Advice, but it remains a challenging area from the perspective of regulatory acceptability. The MHRA team stressed that it’s important that the weight of the external data does not overwhelm the data collected in the trial, but they are open for applications e.g. in the setting of rare diseases. They will also want to see how the data of the new trial looks on its own.