**Reviewer comments: OPTIMUM SAP**

|  |  |  |
| --- | --- | --- |
| **Comment** | **Response** | **Input needed** |
| **REVIEWER 1 (**Eric Arthur Dunipace**): I must say, I am very interested to learn how this trial actually turns out. It seems like an important topic and helping to address one component of the hygiene hypothesis. I also liked the heavy use of Bayesian analysis to answer these important questions. It seems like this makes the maths much easier for calculating superiority/inferiority and answers the the question more intuitively in terms of probability of the parameter being above or below 0.** | | |
| The only thing I thought could be better motivated was the use of q=0.95. It was stated that this was determined through simulations, but it seems suspicious that the simulations just happened to find the common rule of thumb cutoff value...Some justification of why 0.95 was chosen based on those simulations would be helpful. And though I may have missed it, I don't think a reason or criterion was given in the simulation setup for how an optimal value of q would be chosen a priori. | We have included additional motivation for the chosen thresholds in the relevant section.  In summary, our primary aim was to choose thresholds which approximately maintained Type I error of 0.05 under the null scenario (last row in Figures 1 to 4).  Beyond this requirement, we did not specify an objective function for defining optimality, but heuristically investigated the expected sample sizes and marginal probabilities for stopping.  Given that 0.95 is a familiar reference for one-sided testing, we identified for $q=0.95$ the threshold values for futility and expected success $(\underline{c}, \overline{c})$ which approximately maintained the targeted type I error.  For the remaining thresholds, we chose to be conservative in stopping too early for futility (e.g. $(\underline{c}=0.2$, say) or expected success ($\overline{c}=0.8$, say), by requiring reasonably stringent thresholds. | Done |
| You may also consider an adaptive design that updates the assignment probabilities, but maybe this benefit for patients is not worth the extra effort given the outcomes are not terribly severe. | We thank the reviewer for their suggestion and acknowledge this possibility, however, for two arm trials our preference is to maintain equal assignment between arms. | Done |
| **REVIEWER 2 (**Alexander Ooms)**: In general this analysis plan is well laid out and clear. There are some points that need clarification however.** | | |
| Pg 8, line 28. It is unclear where the desired reduction in response rate from 10% to 7% has come from. Is this the MCID? Please justify with reference. | No pre-existing minimum clinically important difference has been established with regard to prevention of food allergy. This MCID I was selected on the basis of consultation with clinical stakeholders. The manuscript has been updated to reflect this. | Done |
| Pg 10, line 49. I'm not sure of the benefit of stating your statistical hypothesis in a frequentist way when you also state a decision rule for trial success. Bayesian hypothesis testing typically relies on Bayes Factor to reflect beliefs about null vs alternative hypothesis which I don't believe you are doing. My understanding is that you won't make statements based on these hypotheses e.g. "reject the null to a x% significance level" so stating these hypotheses in this way arguably does not add value. | The hypothesis as stated is independent of the method of inference and applies equally whether using frequentist or Bayesian methods.  The hypothesis is evaluated by its posterior probability rather than by the associated Bayes factor (a transformation of the hypotheses posterior probabilities).  The posterior probability (and its expected probability of exceeding the specified decision threshold) is used during interim monitoring to inform recommendations relating to futility and expected success of the trial.  When reporting results, the posterior probability of the stated hypothesis will be included along with other posterior summaries from the model. | Done |
| Pg 11, line 29. Study site typically included as a random effect in a mixed effects model, why not here? | We agree that site would typically be included as a random effect rather than fixed effect in a multi-centre trial.  However, due to the small number of planned recruiting sites (3), our preference was to include the sites as fixed effects in the adjusted model. | Done |
| Pg 11, line 53. Perhaps add estimated accrual rate here to give some idea of how many interim analyses may be expected or reference Table 2. | We have included additional information on the expected number of interim analyses that will be conducted based on the expected accrual rate. | Done |
| Pg 13, lines 8-18. Please be more explicit about how many analyses are being performed. Are you doing adjusted & unadjusted analyses on both ITT & PP populations, 4 in total, for each secondary? If so, why so many? If not, clarify what is being done. | The reviewer is correct, there are 4 planned analyses for the outcomes listed in this section: ITT and PP, both unadjusted and adjusted.  The adjusted and unadjusted analyses estimate different quantities (marginal as opposed to conditional) and we see value in reporting both summaries.  Similarly, the ITT and PP analyses target different estimands (intervention policy versus adherence per-protocol), both of which are of interest to the investigators. | Done |
| Is a PP analysis being done for the primary outcome? I could not see this explicitly stated. | We have clarified in the section titled "Analysis of Primary Outcome" that both an ITT and PP analysis will be undertaken for the primary outcome. | Done |
| Pg 13, lines 23-24. Please give these limits. | The seroprotective/seroconversion limits were previously stated on page 6 for each antigen.  We have amended the manuscript to refer back to those limits in the lines referred to by the reviewer.  The limits of quantification have also been outlined in the 6/7. | Done |
| Pg 13, lines 44-45. Is N(0,100) a "flat" enough prior for numbers of this magnitude to be non-informative? | The concentrations and titres will be $\log\_{10}$ transformed, and so we believe the prior variance to be appropriate for the expected outcome scale. | Done |
| Pg 13, line 47. Why use a truncated Cauchy for the standard deviation priors? State the limits set on the distribution to truncate it. Would a gamma be more appropriate? I believe the mean and variance of a Cauchy are undefined so I'm not sure what is written is correct. | The manuscript has been amended to state that these priors will have a lower bound of zero(half-Cauchy).  We have also amended the manuscript to refer to a half-Cauchy distribution with location 0 rather than mean 0 which is undefined as noted by the reviewer, and to refer scale rather than standard deviation which is also undefined.  The prior was selected to be weakly informative on the expected outcome scale (e.g. Gelman (2006)). | Done |
| Pg 13, lines 54-55. Consider a hierarchical structure of visit within patient. Would this look different for Stage 1 and Stage 2 patients due to extra visits in Stage 1? | We thank the reviewer for their suggestion.  Whilst our primary interest is on the marginal effect, we agree that the correlation of repeat visits for the same patient should be accounted for.  We have specified that a random intercept for patient will be included in the model. | Done |
| Pg 13, lines 58-59. State limits. | The seroprotective/seroconversion limits were previously stated on page 6 for each antigen.  We have amended the manuscript to refer back to those limits in the lines referred to by the reviewer.  The limits of quantification have also been outlined in the 6/7. | Done |
| Pg 14, line 38. Close bracket. | This line has been corrected. | Done |
| The simulation study is well reported and described. Consider publishing your code used to perform these simulations as a supplemental material to allow replication of the operating characteristic results and enhance transparency. | We will make the code available in a Github repository which is now linked in the manuscript. | Done |
| **REVIEWER 3 (**Han Zhu)**: This is a good SAP with nice organization. It is good to see SAP using Bayesian adaptive design.** | | |
| (no suggestions) | We thank reviewer 3 for their comments. |  |
| **REVIEWER 4 (LAURA RICHERT): This is a well written paper of the statistical analysis plan of the OPTIMUM study, a randomized double-blind Bayesian adaptative vaccine trial to assess the effects of a wP containing vaccine as the first scheduled pertussis vaccine dose instead of an aP containing vaccine in Australian infants on the risk of IgE-mediated food allergy at the age of 12 months (primary outcome). The design and statistical aspects of this trial are of interest. A paper of the overall trial protocol has been previously published in BMJ Open. In the present manuscript, the authors clearly summarize the statistical aspects and also present the operating characteristics of the Bayesian adaptive design that were assessed by simulation. I only have minor comments that aim to improve the reading of the paper by a reader not yet familiar with the trial:** | | |
| Abstract: unless the journal requirements do not allow for this, I suggest to put a summary of the main operating characteristics of the chosen design in the results section of the abstract. In the current version of the abstract, the true content of the manuscript is only briefly mentioned in the discussion paragraph. | The abstract has been revised to include a results section which briefly summarises the operating characteristics from the trial simulations. | Done |
| Introduction: I suggest to also add a short summary of the immunological rationale for the trial's hypothesis to the first paragraph of the introduction and to cite also the BMJ Open protocol paper here. | We have revised the manuscript to provide a brief summary of the immunological rationale of the trial in the introduction. | Done |
| Page 2, Study synopsis : add a short explanation why the study has two stages (i.e. specific secondary objectives and associated laboratory measurements in the first stage?) | Stage 1 was designed to obtain detailed solicited adverse event data following each primary pertussis vaccine dose, and post-priming immune response data for the first 150 infants only. Stage 2 was designed as a simpler protocol with less intensive follow-up and fewer visits scheduled. | Done |
| Page 4 "Elimination criteria during the study": a more specific title could be chosen for this section (i.e. exclusion criteria from the per-protocol analysis). | We have re-titled this section to reflect it's relationship to the per-protocol population. | Done |
| Page 7, Blinding : mention how the unblinded pharmacist or nurse obtains the next contiguous allocation. Is an electronic tool used for this? | The unblinded pharmacist or research nurse will obtain the next contiguous allocation (i.e. the lowest available randomisation number) from a randomisation list which is concealed in an A4 non-transparent envelope from all other study staff. An electronic copy of the randomisation list is distributed by the study statistician to the unblinded pharmacist for printing and concealment.  The protocol allows for delivery of the IMP/comparator by a blinded or unblinded study nurse depending on site policies- please note that are currently using a unblinded study nurse who is then not involved in any of the follow up for the study.  The manuscript has been amended to provide some additional detail. | Done |
| Define Dk when it is first mentioned. Only ~Dk is defined at the first occurrence (and the definition of Dk (obverved data) is only provided in a later section) | The manuscript has been amended to define $D\_k$ at it's first mention as suggested. | Done |
| Page 12, line 35 : the word 'response' does not seem to be appropriate here, given that the primary outcome is a (undesirable) allergy diagnosis | This line has been rephrased to state "met the primary outcome criteria" rather than "response". | Done |
| Stopping rules : to my understanding, when a stopping rule is met : new enrolments would be halted but the participants already enrolled but with unknown outcomes at the time of the interim analysis would continue follow-up, and the results (including the observed outcomes of these participants) would only be made public when this follow-up is complete. Is this correct? | Yes, this is correct.  The stopping rule refers to terminating enrolment, but follow-up will continue according to the protocol for participants already on study.  On completion, the final analysis will be undertaken and results disseminated. | Done |
| Page 13, line 41 : could a bit more details about the longitudinal model be provided. What random effects would be specified (intercept + time?) | The longitudinal model assumes that a participants responses are multivariate normal with unstructured covariance matrix.  No random effects are specified in the model as interest is primarily in the marginal comparison between intervention groups at each visit. | Done |
| Page 14, line 22 : can a specific rational for the 2/3 non-inferiority margin for the GMR be given? Is the any validated threshold for an immunological correlate of protection for pertussis vaccines? | A non-inferiority margin of 2/3 on the geometric mean concentration ratio is commonly used in vaccine trials and was selected on that basis of available guidance (<https://www.who.int/publications/m/item/WHO-TRS-1004-web-annex-9>). The manuscript has been updated to include this citation.  There are no unequivocal correlates of protection against pertussis. | Done |
| Page 14, line 42: use a separate section title for vaccine satisfaction, which seems to be a concept that is distinct from safety and tolerability | The outcomes relating to vaccine satisfaction have been moved into their own section separate from safety and tolerability. | Done |