Issue 1

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 1.3. Page 10. “First, it is not clear if the concept of a response formally defined by specific changes in only CFS and OSDI is clinically meaningful. If it is accepted that the concept of response is clinically meaningful, then the issue is the lack of evidence available to support the use of any specified threshold value for this measure. The ERG is therefore unable to comment on whether the CFS-OSDI response as defined in SANSIKA (CFS improvement ≥ 2 and OSDI improvement ≥ 23) or the CFS-OSDI response defined post-hoc (also using data from SANSIKA) and used to inform the economic model base case (CFS improvement ≥ 3 and OSDI improvement ≥ 23), is most appropriate.”	Removal from report.	Santen's proposed a composite responder endpoint defining responders as patients at Month 6 with: • A two grade or more improvement from baseline in CFS score on the modified Oxford scale. and • An improvement from baseline in symptom score ≥ 30% (% of baseline score on per patient basis) The proposed composite responder approach consists of a double criteria definition. A sign AND a symptom endpoints are considered corresponding with the objective of defining outcomes measurable for both patients and ophthalmologists. This approach measures simultaneous improvement of both types of endpoints in the same patient and can be viewed as clinically relevant, as supporting treatment rationale and being in agreement with accepted medical practice. The first proposed component of the composite responder analysis is keratitis improvement. Keratitis is a key sign of DED especially in severely affected patients. Keratitis is quantified with a validated tool, the modified Oxford scale. It is considered that a two grade or more improvement of CFS on the modified Oxford scale would represent a	Text not removed. However it is noted by the ERG that its text in 1.3 should state “OSDI improvement ≥ 30%” rather than “OSDI improvement ≥ 23” and so relevant text has been amended. Rationale for response: The ERG notes the additional evidence provided by the company supporting the use for CFS ≥ 2 and OSDI ≥ 30% as valid measures of response. The ERG has not disputed the validity of either CFS or OSDI as individual outcome measures and concurs that a change in CFS is particularly important for measuring impact on keratitis. However, the ERG is not aware of the CFS-OSDI composite response outcome being previously used, or validated, in any other research studies. Nor is the ERG aware of the use of *only* these two measures to be used to assess response in clinical practice. Hence, the ERG considers uncertainty remains as to the clinical meaningfulness of the composite CFS-OSDI endpoint In addition, while the ERG recognises that an improvement of CFS ≥ 3 is more stringent than CFS ≥ 2, the use of

		meaningful change knowing that it is a 0 to 5 points scale. A change of one grade for a patient is clinically meaningful. The second component of the composite responder analysis is improvement of symptoms. The OSDI was selected as a validated patient reported outcome (PRO) instrument in DED. Its reliability, validity and responsiveness have been assessed and established (Schiffman et al. 2000) and this tool has been broadly used in clinical trials over the past few years. It is considered that in moderate to severe dry eye patients, a 30% improvement of OSDI from baseline would be in the range of or above all defined MIDs, and would therefore represent an adequate meaningful clinical difference. The rationale for definition of each response is further explained below. <u>Response on the Sign Component</u> Ocular surface damage is a key sign of DED. Improvement of keratitis/corneal erosion assessed by CFS is a key sign and endpoint for clinical trials in ocular surface inflammatory disease. A two-grade improvement of CFS appears relevant in severe DED patients with CFS grade 4 on the modified Oxford scale at baseline as two grades improvement avoids the issue of the border effect (just above and then just below the class limit), and 2 grades improvement in patients with a score of 4 on the modified Oxford scale represents at least 50% improvement. This is particularly the case in severe patients, such as patients with CFS grade 4, which represent a challenging	this more stringent criteria for CFS-OSDI response is nevertheless a post-hoc analysis; if this is argued to be more clinically meaningful, the ERG queries why this was not used as the primary outcome in the SANSIKA trial. Therefore, in summary, the ERG's two main concerns remain, namely: 1. Is a composite outcome defining response using *only* measures of CFS and OSDI clinically meaningful? 2. If so, what is the rationale for choosing an improvement in CFS ≥ 3 over CFS ≥ 2 for assessing cost-effectiveness when the former was used as part of the primary outcome for assessing clinical effectiveness?

		patient subgroup with eyes at risk for irreversible damage to the ocular surface, and in particular, to the cornea. The clinical relevance of a two grade reduction in corneal staining should stand as clinically meaningful improvement e.g.: • A prevention in visual image degradation if the area involved is in the visual axis • A restoration of tear film stability in DED as the damaged base for the tear film decreases • A reduction in the damaged area associated with the risk of microbial keratitis. In conclusion, the threshold of two grades for an improvement of CFS on the modified Oxford scale defines a relevant response for the sign component in a composite responder endpoint. <u>Response on the Symptom Component</u> The OSDI is a 12-item self-administered PRO instrument. The reliability and validity of the OSDI was investigated by Schiffman et al (2000). The OSDI was found to be valid, effectively discriminating between normal, mild to moderate, and severe dry eye disease as defined by both physician's assessment and a composite disease severity score. The conclusion was that the OSDI is a valid and reliable instrument for measuring the severity of dry eye disease, and it possesses the necessary psychometric properties to be used as an endpoint in clinical trials.	

		As recommended by the FDA PRO guideline (December 2009), the OSDI has been evaluated in terms of reliability, validity and ability to detect change (responsiveness) and has been successfully used in a wide number of clinical trials as an efficacy endpoint over the past few years. The Minimum Clinically Important Difference (MCID) is used to interpret whether the observed change is important from the patient's and/or clinician's perspective. According to PRO guidelines from February 2006, PRO instruments can be useful to specify a MCID as a benchmark for interpreting mean differences. An MCID is usually specific to the population under study and may lead to a definition of a responder. The recommended approach is to estimate the MCID based on several anchor-based methods (see Miller et al., 2010 for OSDI), with relevant clinical or patient-based indicators (e.g. Clinician Global Impression; CGI & Subject Global Assessment; SGA). In the case of the OSDI, Miller and co-workers (2010) presented the MCID for OSDI: The Miller et al article on MCID (Minimal Clinically Important Difference for the Ocular Surface Disease Index, Arch Ophthalmol. 2010; 128(1): 94-101) is a potentially important study which attempts to build on a well-studied database of patients which has been collected by Allergan Inc. in the form of an observational registry compiled between the years 2004-2008, and collected from 75 sites. The principal PRO instrument is the Ocular Surface Disease Index (OSDI) which	

		has been approved by the FDA as a validated measure of patient symptoms for use in clinical trial assessment. The findings here demonstrate that both the CGI and the SGA correlate well with the OSDI and identify MCIDs ranging from 7.3-13.4 along the severity scale in severe dry eye patients. The values corresponding to 30% improvement of OSDI are in the range or even higher than values calculated for MCID by Miller et al., especially for severe dry eye patients (target population). In the SANSIKA study, more than 96% of the patients had an OSDI value at Baseline of 33 or more. A 30% improvement represents an improvement higher than the MCID defined by Miller et al. A ≥ 30% improvement from baseline in symptom score is considered clinically meaningful since higher than the MCID defined by Miller et al. The primary composite endpoint was not met in SANSIKA study however when using a more stringent criterion for the CFS responder rate by increasing the required improvement from at least 2 grades to 3 grades (which is even more clinically relevant), Ikervis® was superior to vehicle at Month 6 (p = 0.016; 18.8% vs. 7.7%). It is important to highlight that the indication is reflecting the benefits of Ikervis® on keratitis (CFS). There was a statistically significant improvement in the CFS score over time in	

		favour of Ikervis®. A decrease of corneal staining was observed in both treatment groups at Month 6 compared to Baseline (-1.76 with Ikervis® and -1.42 with vehicle). The observed difference of 0.35 units between active and vehicle arm appeared rather modest, but when translating the logarithmic scale into actual number of dots of staining, i.e. corneal lesions. The difference represents a ratio of 1.5 in the damaged surface area. This means that the vehicle group presented on average with 50% more dots/lesions compared to the Ikervis® group, which was considered by the CHMP and the experts to be clinically meaningful. (ad hoc expert group convened by CHMP). Santen agrees that SANSIKA study is the more relevant to the disease problem. SANSIKA study was considered as the pivotal study and SICCANOVE study was considered supportive.	

Issue 2

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 1.3. Page 10. "The ERG considers that the value of the evidence from the SANSIKA trial is limited by the fact that it uses the Ikervis® vehicle as the comparator intervention, rather than any of the comparators specified in the NICE scope. Not only is the vehicle not commercially available, it is not currently used in routine clinical practice; in addition, the company argues that it may offer some therapeutic benefit. Certainly, improvements over time were reported for all efficacy outcomes in the vehicle arm of the SANSIKA trial. However, it is not clear whether the improvements occurred as a result of the vehicle, as a result of concomitant AT use, or as a combination of both vehicle and AT."	"The comparison of Ikervis® with its vehicle was considered to be valid by EMA CHMP. Improvements over time were reported for all efficacy outcomes in the vehicle arm of the SANSIKA trial. However, it is not clear whether the improvements occurred as a result of the vehicle, as a result of concomitant AT use, or as a combination of both vehicle and AT."	Santen agrees that the improvement from baseline occurred as a result of the vehicle plus the concomitant AT use. Based on the study design it is not possible to differentiate the vehicle effect alone. The comparison of Ikervis® with its vehicle was considered to be valid by EMA CHMP. A comparison of Ikervis® with other CsA would have been informative however no common comparator was available to conduct such a study.	Text unaltered. Rationale for response: The EMA assesses issues of efficacy and safety but not cost-effectiveness. Clearly the SANSIKA trial demonstrated benefit for Ikervis vs Vehicle, indicating that the active pharmacologic element of Ikervis is incrementally active compared to the vehicle. Establishing the clinical and cost-effectiveness of Ikervis vs currently available clinical comparators is a separate question. For this the NICE requirement of comparison(s) against products representing current clinical practice is required, and the vehicle does not fall within this standard.

Issue 3

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 4.2.4. Page 37. :"The ERG notes that in the company's presentation of the results, the forest plot lacks important detail commonly reported with the presentation of meta-analyses such as the weight given to each study and a test for heterogeneity (such as I2). As such, the ERG has some concerns that the data may have been simply pooled by adding the data together rather than using standard techniques for conducting meta-analyses. This would also mean that the randomisation in the individual studies is unlikely to be preserved."	Amendment to statement on randomisation and weighting.	It should be noted that the meta-analysis was issued from the analysis of the pooling of two phase III studies. This analysis was done at the individual data level (with the initial raw data) which deals both with the issue of weighting the studies according to their size and any randomization concern (which is only relevant when the meta-analysis directly combines different statistical estimators from several studies). In the Meta-Analysis report, a sensitivity analysis is provided including the study*treatment interaction, in order to assess the heterogeneity concern (ICHE9) for all parameters analysed. A copy of this report can be made available on request.	Text deleted and additional text clarifying the analysis was done at the individual data level added. Rationale for response: It was not clear from the company's original submission (nor the company's responses during the clarification process) that the meta-analysis was conducted at the individual data level.

Issue 4

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 5.5.1, page 63. “The results from SANSIKA cannot be used directly to inform an economic evaluation because the comparator (vehicle) is not commercially available”	Amendment to report: Whilst the comparator in SANSIKA is not commercially available, it also has no active pharmaceutical properties. Therefore, the efficacy of the vehicle plus artificial tears has been used in the absence of other available comparators in this space to inform the economic evaluation as a proxy for the efficacy of artificial tears.	We believe the efficacy of vehicle in the SANSIKA trial is a valid source of information for the economic evaluation. Vehicle efficacy has been used as a proxy to parameterise standard artificial tears which are commercially available and included in the NICE scope. The Ikervis® vehicle is compositionally similar to Cationorm® a cationic ophthalmic emulsion which is an efficacious artificial tear commercially available in Europe, though not the UK. The treatment benefit of vehicle in addition to standard artificial tears has been included in the economic evaluation, without attributing the cost of Cationorm®, this conservative assumption is either neutral or biases against Ikervis®. Therefore we believe that lack of commercial availability is not material.	Text unaltered. Rationale for response: Regardless of any assumptions (conservative or otherwise) vehicle cannot be considered a relevant comparator to Ikervis, as it cannot be acquired for use and is not marketed in the UK. As noted in the ERG’s response to Issue 2, establishing the clinical and cost-effectiveness of Ikervis vs currently available clinical comparators is the question at issue and so Section 5.5.1 is accurate.

Issue 5

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 5.5.7. Page 68. Artificial tear use error	Removal from report.	We believe that artificial tear (AT) use for the trial period was applied correctly at the rate shown in the CSR. Six month AT usage was extracted directly from the SANSIKA CSR (page 101, Table 26) which reported drops per eye, per day for both Ikervis® and vehicle. We have assumed that each patient has two affected eyes and the rates of AT usage were multiplied by number of eyes treated to estimate the average drops per day in both treatment arms (F38 to F42 in 'costs and resource use'). However, we have also assumed that single use minims were shared across both eyes (one vial for both eyes), thereby obviating concerns over inflating the cost of AT.	Text to Section 5.5.7 amended and more appropriate AT adjustment incorporated in a revised version of Table 25 (page 74). This has changed the overall ICER and therefore text on pages 15, 69, 72, 73, 78 and 79 (Please also note changes to 10.3 Appendix 3: Implementation of ERG decision model amendments, page 90). Rationale for response: The ERG acknowledges that the amendment applied in the estimation of AT costs is inaccurate. However, the ERG has identified an inconsistency between the calculation of AT use at baseline and at 6 months. The company has argued that there is no basis for distinguishing between the number of drops per eye per day recorded in the two trial arms (16.54 for the vehicle arm and 13.24 in the Ikervis arm) as the difference is not statistically significant. However the difference recorded at 6 months is much smaller (7.32 and 6.34 respectively) and is also not statistically significant. Moreover, the proportionate reduction in AT use is very similar in the two trial arms (55.7% in the vehicle arm and 52.1% in the Ikervis arm). There is therefore no basis for employing different AT use estimates for patients responding to

			treatment in either arm. If an average usage of 6.83 drops per eye per day is applied to the model, the ICER for Ikervis vs vehicle increases to £36,307 per QALY using the SANSIKA protocol definition of response, and to £20,950 per QALY using the company's post hoc definition.

Issue 6

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 5.5.8. Page 68-69. Model coding errors	Removal from report.	Whilst we would be happy to engage in further discourse with the ERG we would respectfully suggest that the structure of the model has been coded correctly but we think there may have been a misunderstanding of the individual time periods the cycles relate to. Specifically, in row 13 of the Ikervis® trace, labelled as cycle 3, 9 months refers to the period 6-9 months. The proportion of patients receiving Ikervis® in this period is modelled as 18.7% reflecting the proportion of patients responding at the end of the period 3-6 months. Therefore, patients are reclassified as responders or non-responders according to trial efficacy assessments, not at the end of the third cycle as the ERG suggest. Calculations (both pre and post half cycle correction) are found in columns AF to AK of the 'Ikervis Trace' and 'Artificial Tears Trace' of the cost-effectiveness model. On reflection we recognise that we could have more accurately labelled cycle 2 as 3-6 months and cycle 3 as 6-9 months which would have avoided this confusion.	Section 5.5.8 removed from report and all reference to coding errors deleted elsewhere in report, i.e. pages 12-13, Rationale for response: The ERG accepts the company's explanation of the aspects of the model logic which were unclear, and concurs that clearer labelling of the model with respect to the distinction between time points and cycle periods would avoid such difficulties.

Issue 7

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 5.5.9. Page 69. Treatment costs	Removal from report.	We believe that not modelling treatment discontinuation should be viewed as conservative as Ikervis is the more expensive treatment.	Text unaltered. Rationale for response: There is no factual error. Clinical advice is that treatment is typically dispensed monthly, and therefore costs should be calculated on the number of patients still on active treatment at the beginning of each month. Whether this change is conservative or otherwise is irrelevant.

Issue 8

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 5.5.10. Page 69. Discounting logic	Removal from report.	Discounting has been applied in the economic evaluation on a quarterly rather than an annual basis. We are not aware of any explicit guidance or economic principle suggesting that discounting should be applied on an annual basis (with no discounting being applied within the first year). However, this does not seem to be a material issue within this assessment.	Text unaltered. Rationale for response: There is no factual error. NHS budgets are set annually, and NHS Reference Costs are calculated annually. Using other than annual discounting over multiple years therefore risks introducing bias into a cost-effectiveness analysis. In this instance the impact of the model amendment is small, and has only a minor effect

In addition, the ERG noticed (and corrected) the following minor errors:

Table 8 (page 41): ‘Change in CFS’ should not have been highlighted

Pages 73 and 74: ‘Table 15’ should be ‘Table 25’

This table 25 was also missing from the List of Tables and has been added here

Consequently, Table 25 on page 86 should now be Table 26

The VBA macro was previously referred to in the report, but the details of the Implementation of ERG decision model amendments were previously omitted and these have now been added at the end of the report (page 92)