Author’s response to reviews

Title: Evaluation of a new disposable silicon limbal relaxing incision knife by experienced users

Authors:

John Albanese (John_Albanese@bd.com)
Geoffrey Dugue (Geoffery_Dugue@bd.com)
Valentin Parvu (Valentine_Parvu@bd.com)
Ann M Bajart (ambajart@eyeboston.com)
Edwin Lee (Edwin_Lee@bd.com)

Version: 2 Date: 9 November 2009

Author’s response to reviews: see over
9 November 2009

Dear Editor in Chief,

Thank you for taking the time to review our manuscript and provide comments. Below each comment is individually addressed. A revised manuscript has also been submitted.

Reviewer's report (6 October 2009):
1. The specification of three knives used in this study must be introduced. And, if possible, the author had better add the qualitative images of them.

Further information regarding the details of the product has been added to the Methods Section of the revised manuscript, as well as an image of each knife has been included in the revised manuscript. The text added to the Methods Section is below:

“The silicon accurate depth knife is a single, uni-directional cutting knife with the depth preset by the manufacturer. The knife also has a retractable shield to protect the user and the blade, and the shield’s slider mechanism is designed to open and close with a single-hand operation. The steel accurate depth knife has the depth also preset by the manufacturer but has bi-directional cutting capabilities. The diamond LRI knife is retractable and can be pre-set to various depths. It is a reusable knife with bi-directional cutting capabilities. Of the three knives the diamond knife is the only one designed for re-use.”

2. In Materials of Methods, last paragraph is repeated twice. (“The dulled steel and silicon knives were ........... dulled on these knives.”)

The repeated text has been removed from the revised manuscript.

3. "Each surgeon made eight incisions with silicon accurate depth knives, eight incisions with diamond LRI knives and two incisions with steel accurate depth knives” --> Why did you make the difference in number of incision among three types of knives?

The difference in the number of incisions in the knife types is related to the hypotheses related to each knife type as compared to the silicon accurate depth knife. The study was
designed to determine non-inferiority between the diamond and silicon accurate depth knife and superiority of the silicon accurate depth knife compared to the steel knife. Based on a pilot study of similar design to this study and our experiences with the devices, it was expected that differences between the steel knife’s performance and the silicon / diamond knives’ performance would be larger than differences between the silicon knife and the diamond knife. For that reason we decided to run more incisions with the diamond and silicon knives than with the steel knife. The pilot study demonstrated a within surgeon coefficient of variation (CV) for the different VAS measurements of between 1% and 30%. The sample size of this study was calculated using an assumed CV of 30%. The assumed CV allowed the superiority hypotheses to have an 86.6% power to conclude superiority if the silicon accurate depth knife was at least 10% better than the steel knife’s performance with the planned number of incisions. The assumed CV allowed the non-inferiority hypothesis to have an 81.3% power to conclude non-inferiority if the silicon knife’s performance was no more than 5% worse than the diamond knife’s performance.

This has been addressed in the revised manuscript by adding the following text to the Methods Section:

“Prior to the study it was expected that differences between the steel knife’s performance and the silicon and diamond knives’ performance would be larger than differences between the silicon knife’s and the diamond knife’s performance. For that reason it was decided to have the surgeons perform more incisions with the diamond and silicon knives than with the steel knife. Based on an assumed Coefficient of Variation of 30% (as observed in a pilot study), if 60 surgeons finished this study, then 480 incisions would have been performed with the diamond knife, 480 incisions with the silicon knife, and 120 incisions with the steel knife. This would have allowed the study to have a 81.3% power to conclude non-inferiority if the silicon knife’s performance was no more than 5% worse than the diamond knife’s performance and a 86.6% power to conclude superiority if the silicon knife’s performance was at least 10% better than the steel knife’s performance in the characteristics measured.”

4. Visual Analog Scale (VAS)
   a) The smoothness of making the incision
   b) How well the curvature of the eye was tracked
   c) The control of the incision
   d) The overall incision quality
   e) The sharpness of the blade
   --> The meaning of each questionnaire may be vague to the individuals. How can the participants estimated the sharpness of the blade? Grossly? It is better that the more defined informations for a VAS are added.

The authors acknowledge the potential vagueness of the descriptors to any given individual. However, as these aspects of performance are very subjective, we believe vague descriptors are necessary to appropriately capture the human perception of using
these devices. For example, even though the physical dimensions of and forces (such as drag force) applicable to the devices could be measured in a laboratory, it would be unknown if these factors actually affect the human perception of sharpness or the other characteristics assessed. Further studies may need to be conducted to determine if these factors relate to improved clinical outcomes, but this study aids in laying the groundwork.

This has been addressed in the revised manuscript by adding the following text to the Discussion Section:

“There is currently no ‘gold standard’ objective definition of sharpness. Therefore, we decided that an evaluation of the subjective experience of highly qualified users would be the best way to assess the differences between these devices. There is also no accepted scale or standard descriptors to measure sharpness (or any of the characteristics assessed), hence we believe the descriptive terms we used, validated as they were within the study, are the best tool we have to capture human perception of the sharpness of these devices. Even though precise measurements may be performed in a laboratory setting there is not sufficient research to link measurements (such as drag force or blade architecture) to a conclusion of which device surgeons would prefer. The descriptors we used in a standard VAS format were able to repeatably discriminate differences between test groups in a predictable, logical direction. For example, scores on the scales for diamond knives that were purposely dulled were found to be grouped near each other and these values were significantly lower than scores for identical knives which were not dulled.”

**Reviewer's report (17 July 2009):**
The authors present a well-designed study which compares silicon limbal relaxing incision knife with diamond and steel knives.

A few specific comments may contribute to the manuscript:

1. **Page 3: second paragraph in materials repeated**

   The repeated text has been removed from the revised manuscript.

2. **Why surgeons did 8 incisions with three types of knives, and only 2 with steel accurate depth knives?**

   Please refer to response 3 above.

3. **Justify the use of a general linear model and Tukey’s HSD multiple comparisons in the methods section**

   A general linear model with surgeon included as an effect is needed in fitting this type of data because there are differences in how different surgeons (subjects) fill out a Visual Analog Scale. Some surgeons may rate all blades on the high end of the scale, while others may rate them all on the lower end, depending on their prior personal experience.
The need for multiple comparison adjustments is somewhat debatable. In general we prefer to use these adjustments because it makes results be conservative. For instance in our study, by not using multiple comparisons and concluding significant differences and/or equivalence between devices, one may wonder what is the overall error rate. If each comparison is done at a 5% significance level, then there is potentially as high as 14% risk that at least one of the conclusions is wrong. By using Tukey’s adjustment we ensure that there is only a 5% overall risk of making an incorrect claim. From a general regulatory standpoint as well, the FDA usually prefers multiple comparison adjustments to be made in these types of analyses. The reasoning is that you may run for instance your device versus a large number of competitor / predicate devices and without adjustments you may conclude superiority versus one of them purely by random chance.

This question has been addressed in the revised manuscript by:

Moving the following text from the Results to the Methods Section: “The VAS measurements of the non-dulled knife ratings were log-transformed and a general linear model with surgeon as a random effect was fitted for each characteristic. Pairwise comparisons using Tukey’s adjustment for multiple comparisons were made among the three different knife types.”

And by adding the following text to the methods section: “VAS measurements were log-transformed in order to normalize the residuals and stabilize the variance.”

3. Discussion: information in the last two paragraphs pertains to methods section, at least in part. However, limitations and strengths of the study are not sufficiently discussed

The authors have further addressed the limitations and strengths of this study in the Discussions Section of the manuscript. The added and revised portion of the Discussion Section is below:

“To further limit any potential bias in this study, a secluded booth at the far edge of the New Orleans Convention Center exhibit hall was used, completely independent of the commercial BD booth. The physical distance aided in ensuring that the study participants consisted of a wide variety of surgeons who visited the trade show floor rather than merely those who wished to visit the BD booth. The study was conducted by the BD Clinical Trial Resources Department and all study conduct was monitored under strict GCP (Good Clinical Practices) guidelines. The study staff was trained to not discuss design of the devices being evaluated or any topics related to BD or BD products in order to prevent their opinions from influencing the surgeon’s ratings. Only study staff and the participating ophthalmic surgeons were allowed into the booth to limit any influence that sales or marketing personnel may have had on the evaluations. The study population was also representative of those who would use the products for surgery. As displayed in the study results, the surgeons who participated in this trial had an average of 15.4 years of experience, an average frequency of 28.9 limbal relaxing incisions performed in the 6
months prior to the study and a roughly even division of diamond/steel knife users. Also, all Visual Analog Scale measurements were made by an optical scanner to reduce measurement error and bias. The order of the evaluations was also randomized in an effort to eliminate any bias that could have occurred due to rating the knives in a consistent order. Randomization also should have reduced the effect of any unknown variables on the outcome measures.

Even though this study was designed to overcome many areas that may have introduced bias, the study is not without limitations. First, the experience of performing an incision in the *ex-vivo* porcine eye may not be representative of an incision made in a human eye. The difference between live human eyes and an *ex-vivo* porcine may impact a surgeon’s perception of a device. However, since each device was compared in the same model, these relative performance assessments should hold true in most applications. Further, porcine eyes have been shown to provide reliable data analogous to that from humans in many areas of Ophthalmology. As described, steps were taken to standardize the model through controlling the age of the eyes following harvest. Eye pressure and eye moistness were also maintained in a standardized fashion across all eyes. Eye pressure control was provided by the use of suction on the back of the porcine eye to simulate intraocular pressure. The same stand also provided stable fixation of the eye. Appropriate moisture on the cornea was maintained by frequent spraying of balanced salt solution on the eyes throughout the evaluations. While no model can fully represent the condition of surgery on living, human tissue, our intent was to provide the most accurate model possible.

We could not assess whether important clinical outcomes such as wound healing and wound stability would be improved by the use of instruments surgeons perceive to have better performance characteristics, such as being sharper. However, to the extent that sharpness may be correlated with better operator performance, we believe that the clinical outcomes found when using the silicon knife would be equivalent to that of the diamond knife. Furthermore, one may consider that patient safety may well be improved with the decreased risk of patient to patient contamination resulting from utilization of a single use device.

There is currently no ‘gold standard’ objective definition of sharpness. Therefore, we decided that an evaluation of the subjective experience of highly qualified users would be the best way to assess the differences between these devices. There is also no accepted scale or standard descriptors to measure sharpness (or any of the characteristics assessed), hence we believe the descriptive terms we used, validated as they were within the study, are the best tool we have to capture human perception of the sharpness of these devices. Even though precise measurements may be performed in a laboratory setting there is not sufficient research to link measurements (such as drag for or blade architecture) to a conclusion of which device surgeons would prefer. The descriptors we used in a standard VAS format were able to repeatably discriminate differences between test groups in a predictable, logical direction. For example, scores on the scales for diamond knives that were purposely dulled were found to be grouped near each other and these values were significantly lower than scores for identical knives which were not dulled. Overall, even though this study does not answer all the questions in regard to which knife to choose...
when making a limbal relaxing incision, it does provide a reasonable approach for future evaluations of LRI knives."

Thank you for continuing to consider this article for publication. We look forward to hearing from you.

Sincerely,

John Albanese, MBA, CCRP

Corresponding Author Information:

John Albanese, MBA, CCRP
Becton Dickinson and Company
1 Becton Drive, Franklin Lakes, NJ 07417 USA
Ph: 201-847-6764
Fax: 201-847-5213
Email: John_Albanese@bd.com