Journal of Advances in Medical and Pharmaceutical Sciences
1(1): 30-39, 2014

SCIENCEDOMAIN *international*
www.sciencedomain.org

Reproducible, Repeatable and Clinically-relevant Hemostasis Scoring

Julianne M. Siegel[1], John F. Cummings[1] and Jeffrey W. Clymer[1*]

[1]Ethicon Endo-Surgery, Inc., Cincinnati OH, Ohio, United States.

Authors' contributions

This work was carried out in collaboration between all authors. Authors JMS and JFC contributed to the protocol, study execution, analysis, and writing of the paper. Author JWC contributed to the analysis and writing of the paper. All authors read and approved the final manuscript.

Method Article

Received 27th May 2014
Accepted 1st August 2014
Published 16th August 2014

ABSTRACT

Background: Although the assessment of hemostasis is frequently performed during stapling procedures, and is necessary during product development, there is currently no widely-accepted scoring system for its evaluation. In order to better measure stapler performance, we developed a 5-point Likert-type scale to evaluate intra-operative bleeding in stapling applications.
Methods: The scale was evaluated internally by novice and experienced appraisers, and then used by 154 surgeons that viewed video segments of stapling in pig that displayed varying levels of hemostasis.
Results: Reproducibility and repeatability of the scale among internal appraisers, as well as the surgeons, was excellent with Kendall's W of 0.94-0.98 (repeatability) and 0.93-0.95 (reproducibility) internally and 0.85 and 0.95, respectively for surgeons. More than 90% of surgeons felt the hemostasis scale was clinically relevant for evaluating hemostasis during stapling.
Conclusions: The scale should be a useful tool in assessing hemostasis with high confidence in its repeatability and reproducibility.

Keywords: Hemostasis; stapling; score.

**Corresponding author: Email: jclymer@its.jnj.com;*

1. INTRODUCTION

Laparoscopic surgery has become highly dependent upon the use of staplers based on their ability to consistently produce quick and durable tissue appositions and anastomoses. While the most devastating complication from staple line failure is leakage, intra-operative bleeding can also be a substantial problem occasionally requiring transfusions or conversion to an open procedure. A small amount of bleeding at the staple line can cause obstruction of the view and alter the rhythm of the procedure, leading to prolonged operating times. Alternate hemostasis techniques, such as cauterization or clips may be necessary, increasing both the time and costs of surgery.

To reduce the incidence of staple line leaking and bleeding, various techniques have been introduced, such as the use of buttress reinforcement material [1], addition of staple lines and alteration of their geometry [2], oversewing of the staple line [3], and modification of staple height [4]. No standard method of measuring bleeding from staple lines has been agreed upon yet; the list of assays includes simple observation of presence/absence of blood, visual blood volume estimation, gravimetry, and measuring time until bleeding stops [5].

Comparison of bleeding from different stapling devices and techniques requires a metric that ideally should be easy, fast, repeatable among and reproducible between observers, and clinically relevant. With these in mind, we have developed a 5-point Likert-type scale for hemostasis. A subjective assessment like the Likert-type scale can be easy and fast if the scale is designed simply and intelligibly. In this study we sought to determine whether our hemostasis scale was also repeatable and reproducible when in use by internal expert and novice appraisers. We then confirmed that ratings could be performed by actual surgeons, and whether results could be linked in a meaningful way to clinical applications.

2. METHODOLOGY

All procedures described in this study were covered by protocols reviewed and approved by the Ethicon Endo-Surgery, Inc. Institutional Animal Care and Use Committee, in accordance with the requirements of the United States Animal Welfare Act.

Video recordings were made of stapling in porcine gastroepiploic pedicles with cartridges of varying staple heights (Fig. 1a-c). The procedures were performed open, but the video recordings simulated laparoscopic procedures. The animals used had no known platelet dysfunctions or coagulation factor deficiencies. A selection of the videos that exhibited a wide range of hemostasis as a result of varying staple height and tissue thickness was used in assessing a Likert-type scale [6] of hemostasis (Table 1). This study consisted of internal experienced and novice appraisers, and external appraisers (practicing surgeons) rating the hemostasis of the recordings.

2.1 Internal Assessment

The internal novice appraisers were veterinary technicians with experience in surgery, but limited experience with the hemostasis rating scale. The internal experienced appraisers had both an extensive background in surgery (>10 years), and were experienced with the hemostasis rating scale and its development. For internal assessments, video recordings of 30 s duration were used to evaluate reproducibility (comparing ratings between appraisers

after watching the video once) and repeatability (comparing ratings within one appraiser after watching the video twice). Viewings that were repeated were performed on different days. The appraisers were blinded to the cartridges used to create the firing, which were performed by technicians not involved in the study. Appraisers were considered either experienced or novice depending upon their familiarity with stapling and hemostasis evaluation. Three simulated studies were created with 26 videos performed with internal experienced observers. Three studies with 52 videos (26 videos repeated twice) were performed with internal novice observers.

Table 1. Hemostasis rating scale

Score	Description
1	No bleeding at tissue site after initial blotting of staple line
2	Blood oozing at tissue site; stops prior to 15 seconds; no intervention needed
3	Blood oozing, still progressive after 15 sec., no intervention needed
4	Blood oozing at tissue site, mild intervention (i.e., cautery)
5	Significant bleeding requiring intervention such as extensive coagulation or ligation with clips

2.2 External Assessment

An online approach was used to conduct interviews among surgeons externally. Interviews were conducted with a target of 150 surgeons, representing bariatric, thoracic, general, colorectal and urological surgeons. All surgeons must have been in practice for 2-30 years, board-certified in general surgery, and currently performing at least 50% of their procedures endoscopically.

Each surgeon evaluated hemostasis in 15 video clips of 10 seconds each. The surgeons were blinded to the cartridges used to create the firing. There were 10 unique videos. In order to determine repeatability, five of the videos were replayed to each surgeon. Due to survey length, the surgeons were broken into two groups. The first group viewed the 10 original clips and 5 duplicate clips (half of the original series), and the second group viewed the same 10 original clips, but a different set of 5 duplicate clips (the half of the original series that the first group did not see a second time). All viewings for each surgeon were performed on the same day.

The length of the videos was chosen to simulate the time constraints of actual surgical procedures. The number of videos and appraisers was chosen based on pilot studies of non-recorded procedures, and consideration of the limit of attention span of viewers. For both internal and external studies, agreement was assessed both between and within appraisers via Kendall's coefficient of concordance (Kendall's W) [7]. Kendall's coefficient was used to assess the Likert-type scale because this statistic evaluates agreement with appraisers' rating weighted by the degree of difference; misclassifying a '5' as a '1' is worse than misclassifying it as a '3'. A Kendall's coefficient of 0.70 or greater was considered to be evidence of a scoring system with appreciable concordance, and a coefficient of 0.90 or greater was considered to be excellent concordance [8].

After viewing the videos and rating hemostasis using the scale, surgeons were asked to evaluate the ability of the scale to discriminate levels of hemostasis, whether the scale was

clinically relevant, and what hemostasis score represents minimally-acceptable performance for a stapler.

3. RESULTS

3.1 Internal Assessment

Kendall's Coefficients of Concordance were above 0.90 both when calculated between and within appraisers, for both experienced and novice appraisers (Table 2) showing excellent reproducibility and repeatability, with significant concordance in all cases.

Table 2. Kendall's coefficient of concordance for internal assessment

Appraiser base	Between appraiser	Within appraiser
Experienced (n=2)	0.95 (P<.001)	0.95-0.98 (P<.001)
Novice (n=6)	0.93 (P<.001)	0.94-0.95 (P<.001)

3.2 External Assessment

The overall Kendall's coefficient of concordance between appraisers was 0.848, showing acceptable reproducibility between scorers (Table 3). All surgical groups had a coefficient greater than 0.70, and except for Urologic, greater than 0.80 (Fig. 2). The average Kendall's coefficient of concordance within appraisers was 0.95, showing excellent repeatability. All surgical groups had a mean coefficient greater than 0.90 (Fig. 3).

Table 3. Kendall's coefficients for between and within appraisers for individual groups and overall

Group	Between appraisers (10 unique videos)	Within appraiser (St Dev) (5 repeated videos)
Bariatric (n=30)	0.851 (p<0.001)	0.937 (±0.102)
Colorectal (n=45)	0.867 (p<0.001)	0.958 (±0.076)
General (n=19)	0.897 (p<0.001)	0.983 (±0.024)
Thoracic (n=30)	0.873 (p<0.001)	0.955 (±0.045)
Urologic (n=30)	0.782 (p<0.001)	0.926 (±0.153)
Overall (n=154)	0.848 (p<0.001)	0.950 (±0.094)

Surgeons tended to use the extreme ends of the rating scale, with "3" (bleeding progressive after 15 sec, no intervention needed) the least frequently used scale point. Post-use, 90% of surgeon felt that the proposed scale adequately discriminates levels of hemostasis, and 92% felt that the scale is clinically relevant. A hemostasis score of 3 or lower was thought to represent the minimally-acceptable level of performance by 85% of the surgeons (see Table 4).

Table 4. Surgeons identifying the particular level of hemostasis as the minimally-acceptable performance from a stapler

Hemostasis score	1	2	3	4	5
Surgeons identifying the level as minimally-acceptable performance	4%	45%	36%	12%	3%

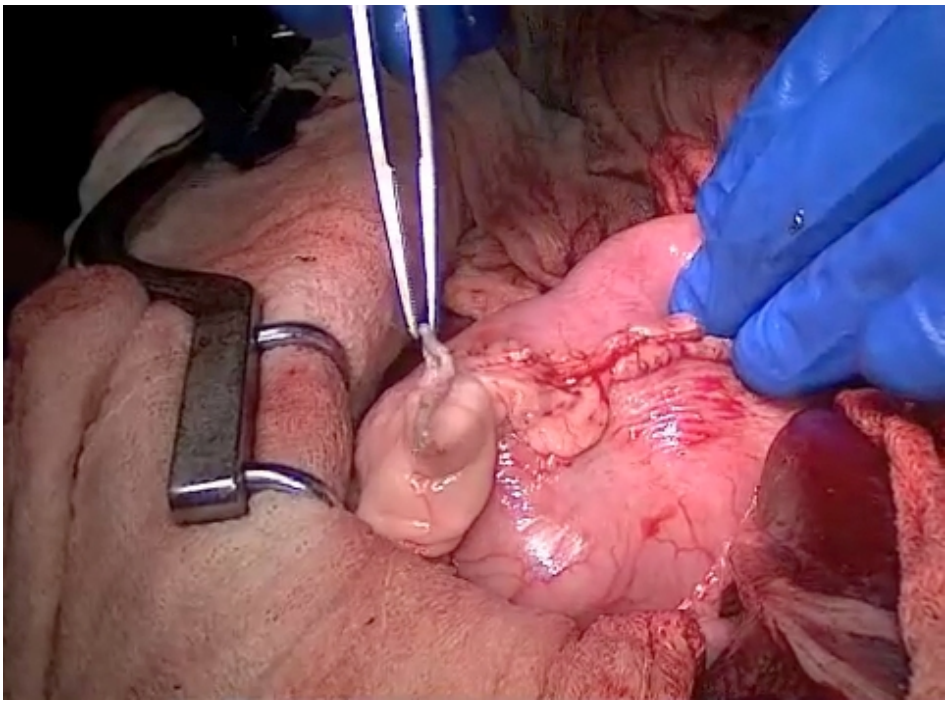

Fig. 1a. Video frame from a stapling with an expert rating of "1" for hemostasis

Fig. 1b. Video frame from a stapling with an expert rating of "3" for hemostasis

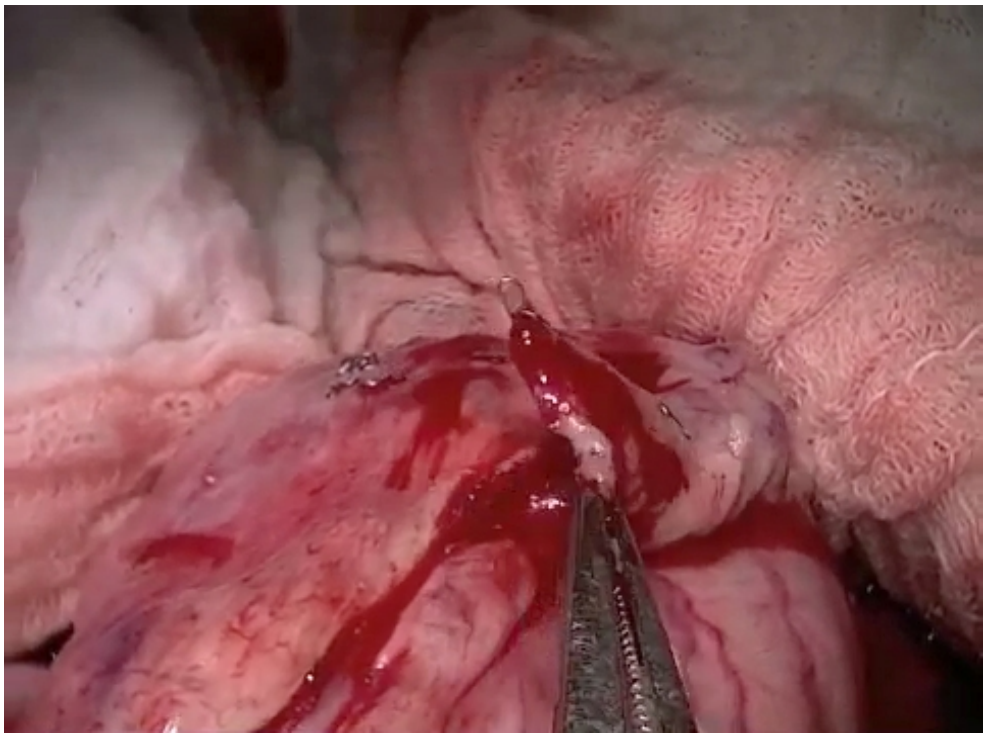

Fig. 1c. Video frame from a stapling with an expert rating of "5" for hemostasis

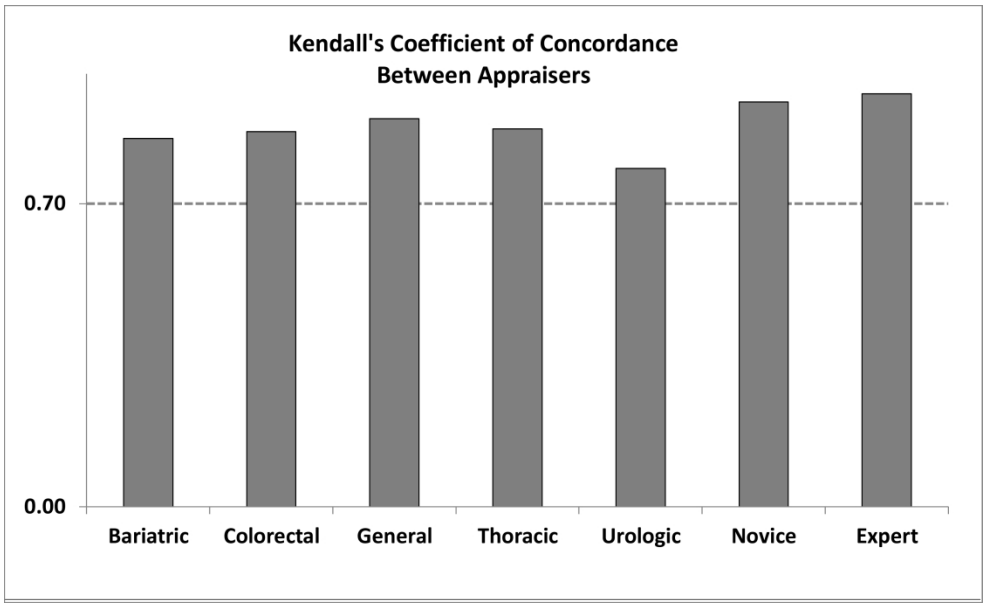

Fig. 2. Kendall's coefficient of concordance between appraisers for surgical groups, novice and experienced internal appraisers

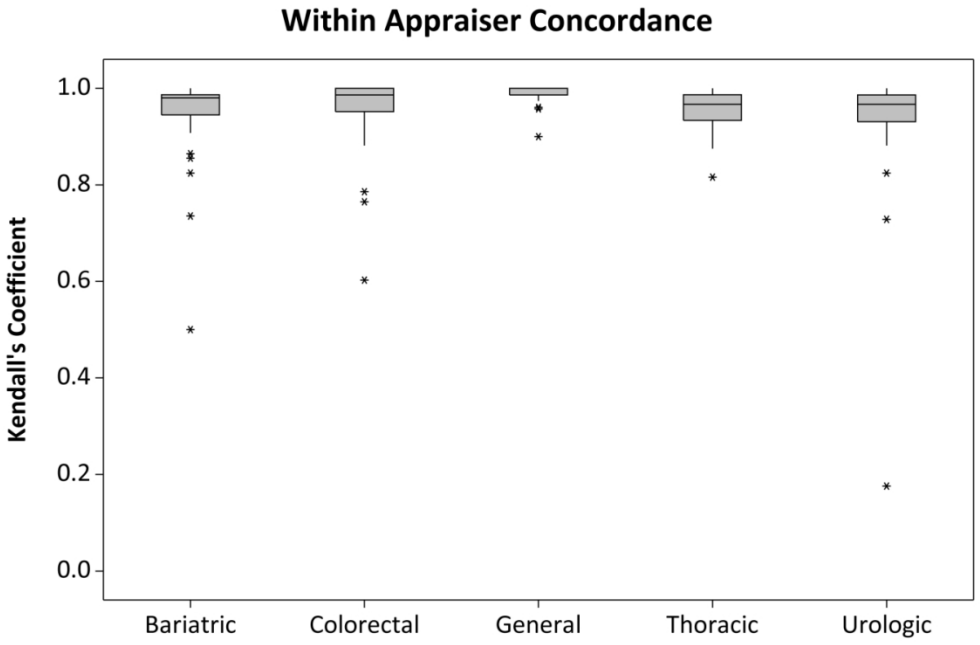

Fig. 3. Kendall's coefficient of concordance within appraisers for surgical groups

4. DISCUSSION

Various methods have been used to describe hemostasis of staple lines intra-operatively, including visually estimating volume, collecting and weighing blood, simply noting the presence or absence of bleeding, and measuring the time for bleeding to stop.

No standard practice has emerged and all these methods have their drawbacks. Quantitative methods, such as gravimetrically assaying blood amount or measuring bleeding time, would be difficult to perform during surgery, and lack clinical relevance when evaluating a stapling device. Qualitative methods, such as visually estimating volume, are prone to inaccuracy, and similarly to merely observing the presence or absence of bleeding, may not provide enough information to make clinically-relevant decisions. A subjective Likert-type scale for hemostasis, while not based on a quantitative measure, may provide a more clinically-relevant method of evaluating stapler hemostasis, in addition to being quick and easy to perform.

No hemostasis score is widely used yet for assessing bleeding during or after stapling. In a study of an energized vessel sealing device applied during partial nephrectomy [9], a 5-point scale was developed similar to the one used in this study, although with the order reversed. Bleeding was graded on a scale of 0=no hemostasis; 1=steady bleeding; 2=moderate bleeding; 3=parenchymal oozing; and 4=dry. For that scale, bleeding scores of 2 or less were considered clinically significant, requiring supplemental coagulation. The authors noted that the amount of bleeding is difficult to estimate during laparoscopic surgery. Their scale was not designed for estimating blood loss, but rather as a clinically relevant instrument. A low score (poor hemostasis) would indicate the need for conversion from laparoscopic to

open surgery. Although their scale is subjective, they note that the grading is performed by the same surgeons who would be making the judgment call on conversion.

The scale in the energized device hemostasis study was presented without any indication of validation. In the absence of a universally agreed upon measure of hemostasis, we sought to determine whether a similar Likert-type scale could be used to assess hemostasis in stapling applications, and whether such a scale is repeatable and reproducible. Our internal results indicated the grading according to the scale is repeatable and reproducible with both experienced and novice appraisers. Results among actual practicing surgeons also showed good agreement between and within surgeons. Of the 154 surgeons that viewed the videos, only one (a urologist) appeared to have substantial problems in using the scale. Surgeons from different specialties all had similarly good results, with only the urology group slightly different based on the one aforementioned participant. A limitation of the current study is that the internal appraisers were all from the same institution, and so the use of Kendall's coefficient, which requires subjects from different institutions, is not appropriate. However, the external appraisers, whose results were similar to those of the internal appraisers, came from a wide variety of hospitals, providing support for the use of the Kendall statistic.

After use, 90% of the surgeons felt that this 5-point scale was sufficiently discriminating of hemostasis; among the minority that was dissatisfied there was no consensus as to a preference for more or fewer levels. A similarly high number of surgeons felt the scale was clinically relevant, lending support to the practice of using these ratings to evaluate the performance of new stapling devices and procedures. Since surgeons tended to underuse the middle of the scale (a score of "3"), application of Kendall's coefficient, which is sensitive to the order of the scale, was confirmed to be the appropriate statistic for this study. Surgeons may have avoided use of the middle of the scale because of the limited length of time of the videos. Interestingly, while the majority of surgeons felt that a hemostasis score of two to three (requiring no hemostatic intervention) was minimally-acceptable, there was a contingent who seemed willing to tolerate the use of stapling devices that produced a score of four (requiring mild intervention).

This study was limited by the use of video clips that were only 10 seconds long (for the external surgeon component), even though the hemostasis scale references bleeding times of 15 seconds. Surgeons did not object to making assessments within this limitation, based on their experience in estimating the amount of bleeding and whether bleeding would stop in a short time. The study required that we show recorded procedures rather than live, but since the objective was to assess a bleeding score in laparoscopic stapling applications, viewing of recording on a monitor did not differ greatly from actual operating conditions. Although viewing videos may limit ability to rate hemostasis, they allowed an assessment of repeatability that live firings could not. While this study was performed in pigs, not in humans, our own experience is that tissue stapling and the appearance of bleeding is similar between the species.

Although the Likert-type scale has been demonstrated to be repeatable, reproducible, and clinically relevant, it is not quantitative. Since gravimetric tests that we have developed separately are limited to thin tissue, such as thin mesentery, the Likert-type scale is more general applicable, as it can be used over a wide range of tissue thicknesses. The limitation with the gravimetric method is that the inherent variability in the animal model prevents an absolute value for hemostasis from being determined, so the stapling evaluations are always performed as relative evaluations, comparing a test device to a predicate.

5. CONCLUSION

We have developed a hemostatic score for use in stapling applications that shows good reproducibility and repeatability with internal associates as well as among actual practicing surgeons. Practicing surgeons agreed that the scale is clinically-relevant and adequately discriminates among the different types of bleeding they typically encounter during stapling procedures. The hemostatic scale should be an excellent tool in the evaluation and comparison of new stapling devices and procedures.

CONSENT

Not applicable.

ETHICAL APPROVAL

All authors hereby declare that "Principles of laboratory animal care" (NIH publication No. 85-23, revised 1985) were followed, as well as specific national laws where applicable. All experiments have been examined and approved by the appropriate ethics committee

ACKNOWLEDGEMENTS

This study was funded by Ethicon, Inc., a Johnson & Johnson company, that manufactures devices used in minimally-invasive procedures.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Consten EC, Gagner M, Pomp A, Inabnet WB. Decreased bleeding after laparoscopic sleeve gastrectomy with or without duodenal switch for morbid obesity using a stapled buttressed absorbable polymer membrane. Obes Surg. 2004;14:1360-6.
2. Sozutek A, Colak T, Dag A, Olmez T. Comparison of standard 4-row versus 6-row 3-D linear cutter stapler in creation of gastrointestinal system anastomoses: A prospective randomized trial. Clinics (Sao Paulo). 2012;67:1035-8.
3. Musella M, Milone M, Bellini M, Leongito M, Guarino R, Milone F. Laparoscopic sleeve gastrectomy. Do we need to oversew the staple line? Ann Ital Chir. 2011;82:273-7.
4. Sakran N, Assalia A, Sternberg A, Kluger Y, Troitsa A, Brauner E, Van Cauwenberge S, De Visschere M, Dillemans B. Smaller staple height for circular stapled gastrojejunostomy in laparoscopic gastric bypass: Early results in 1,074 morbidly obese patients. Obes Surg. 2011;21:238-43.
5. Spector D, Perry Z, Konobeck T, Mooradian D, Shikora S. Comparison of hemostatic properties between collagen and synthetic buttress materials used in staple line reinforcement in a swine splenic hemorrhage model. Surg Endosc. 2011;25:1148-52.
6. Likert R. A technique for the measurement of attitudes. Archives of Psychology. 1932;140:1-55.
7. Siegel S, Castellan NJ. Jr. Nonparametric Statistics for the Behavioral Sciences, Second Edition. McGraw-Hill. 1988;262-270.

8. Lesaffre AE. Statistical and methodological aspects of oral health research. John Wiley & Sons Ltd, West Sussex UK; 2009.
9. Jackman SV, Cadeddu JA, Chen RN, Micali S, Bishoff JT, Lee BR, Moore RG, Kavoussi LR. Utility of the harmonic scalpel for laparoscopic partial nephrectomy. J Endourol. 1998;12:441-4.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=577&id=36&aid=5753