Systematic Review on Mentoring and Simulation in Laparoscopic Colorectal Surgery

Danilo Miskovic, MD, FRCS, Susannah M. Wyles, MSc, MRCS, Melody Ni, MSc, Ara W. Darzi, FMedSci HonFREng KBE, and George B. Hanna, PhD, FRCS

Objective: To identify and evaluate the influence of mentoring and simulated training in laparoscopic colorectal surgery (LCS) and define the key components for learning advanced technical skills.

Background: Laparoscopic colorectal surgery is a complex procedure, often being self-taught by senior surgeons. Educational issues such as inadequate training facilities or a shortfall of training fellowships may result in a slow uptake of LCS. The effectiveness of mentored and simulated training, however, remains unclear.

Methods: We conducted a systematic search, using Ovid databases. Four study categories were identified: mentored versus nonmentored cases, training case selection, simulation, and assessment. We performed a meta-analysis and a mixed model regression on the difference of the main outcome measures (conversion rates, morbidity, and mortality) for mentored trainees and expert surgeons. We also compared conversion rates of mentored and nonmentored.

Meta-analysis of risk factors for conversion was performed using published and unpublished data sets requested from various investigators. For studies on simulation, we compared scores of surveys on the perception of different training courses.

Results: Thirty-seven studies were included. Pooled weighted outcomes of mentor cases (n = 751) showed a lower conversion rate (13.3% vs 20.5%, \( P < 0.0001 \)) compared with nonmentor cases (n = 695). Compared to expert case series (n = 5313), there was no difference in conversion (\( P = 0.2835 \)), anastomotic leak (\( P = 0.8342 \)), reoperation (\( P = 0.5680 \)). A meta-analysis of training case selection data (n = 4444) revealed male sex (\( P < 0.0001 \)), previous abdominal surgery (\( P = 0.0200 \)), a BMI greater than 30 (\( P = 0.0050 \)), an ASA of less than 2 (\( P < 0.0001 \)), colorectal cancer (\( P < 0.0001 \)) and intra-abdominal fistula (\( P < 0.0001 \)), but not older than 64 years (\( P = 0.4800 \)), to significantly increase conversion risk. Participants on cadaveric courses were highly satisfied with the teaching value yet trainees on an animal course gave less positive feedback. Structured assessment for LCS has been partially implemented.

Conclusion: This review and meta-analysis supports evidence that trainees can obtain similar clinical results like expert surgeons in laparoscopic colorectal surgery if supervised by an experienced trainer. Cadaveric models currently provide the best value for training in a simulated environment. There remains a need for further research into technical skills assessment and the educational value of simulated training.

(Ann Surg 2010;252:943–951)

Evidence exists for lengthy proficiency gain curves for technically complex interventional procedures. In laparoscopic colorectal surgery (LCS), this is coupled with a limited dissemination of the technique despite the supportive evidence from randomized controlled clinical trials. In 2006, 32.6% of all colorectal resections in the United States, and 10.4% in the United Kingdom, were performed laparoscopically. Among other factors, the limited uptake of LCS may also be attributed to educational issues such as inadequate training facilities or a shortfall of fellowships and other training posts. Furthermore, at present only limited structured guidelines exist for the training of LCS. The increasing emphasis on patient safety and governance issues necessitate a prudent uptake of technically complex procedures and the framework for training in this context remains unclear. In most Western countries, structured curricula have been set up for training basic laparoscopic procedures, yet there are few reports on dedicated programs for advanced techniques. LCS is a good example of a common complex procedure that is technology-dependent, often being learnt by senior trainees, and has been considered as an index procedure for advanced skills training.

The aim of this systematic review was to identify and evaluate the influence of mentoring and simulated training in LCS and delineate the key components for learning advanced technical skills without compromising patient safety.

METHODS

Question and Search Strategy

We searched Ovid MEDLINE (1950–2009, August Week 3), EMBASE (1980–2009 Week 34), and PsycINFO (1967–2009, August Week 3). We used 3 different domains of MeSH-terms and key words combined by “AND,” and in each domain the terms were combined by “OR.” The first domain contained terms on training, the second on surgical anatomy (colon and rectum), and the third on laparoscopy (Supplemental Digital Content 1). The selection was not restricted by language. Two investigators reviewed titles and abstracts resulting from the search, and those clearly unrelated to the topic and any duplicates were excluded. For those possibly fulfilling inclusion criteria, full text was retrieved and assessed for eligibility. At this stage, one article was identified through cross-referencing and included for final evaluation.

Eligibility and Study Categorization

On reviewing the selected articles, 4 main categories were identified. The first included reports on patient outcome data from mentored and nonmentored training surgeons. Mentored, in this context, referred to whether an expert laparoscopic colorectal surgeon provided technical support, advice, and guidance to the trainee during the operation, either by actually being present in the operating room or remotely via video links (telementoring). We defined the first 50 cases as representative of the early learning curve to be consistent with studies demonstrating a plateau in the learning curve at this level of experience.

The second category was training case selection. These studies assessed certain patient attributes to determine factors that might
predict a more difficult operation and thus increase the likelihood of conversion from a laparoscopic to an open procedure. The third category was simulation, which included studies on different simulated training techniques. The fourth, skills assessment, incorporated articles on assessment methods used to describe the quality of technical performance of LCS trainees.

We excluded case reports, letters or comments, editorials, bulletins, reviews, descriptions of techniques, and studies on hand-assisted, open, or robotic surgery, and reports with no extractable data. If there were incomplete published data, we asked the corresponding author for this by e-mail. If investigators failed to respond to a repeat request or were unable to provide the desirable data, their study was excluded. The search was performed with a good interrater reliability (Cohen’s $\kappa = 0.77$).

**Data Extraction, Outcome Measures, and Analysis**

Two reviewers applied inclusion criteria independently and extracted data into a standardized electronic format (Excel 2004 for Mac, Microsoft Corporation, Redmond, WA), guided by the Cochrane Reviewers’ Handbook.22 Included articles were first allocated to the 4 groups specified within the study objectives, and then relevant data were extracted according to the category.

The outcome measures were determined by conventional clinical outcomes and desirable criteria for simulation and assessment.18,23–25 The end points for each category were as follows:

1. **Mentored versus nonmentored**: clinical outcome (morbidity, mortality, conversion from laparoscopic to open procedure);
2. **Training case selection**: clinical outcome (morbidity, operation time, mortality, conversion from laparoscopic to open procedure);
3. **Simulation**: fidelity, effect on clinical training, trainee satisfaction.18 (Fidelity in this context is defined as the degree to which a simulation matches the real system and/or the environment in terms of physical and functional characteristics.)26
4. **Skills assessment**: reliability, validity and feasibility, acceptability, educational impact, cost-effectiveness.23–25

Two authors extracted conversion rates and morbidity data from reports on 3 subgroups; mentored, and nonmentored cases, and expert cases. The data from the expert series were taken as a control.
For the analysis, we used the software Comprehensive Meta-Analysis (Version 2.0, Biostat, Englewood, NJ). We performed a meta-analysis for each subgroup separately, using a fixed and random effects model for event rates. We did a subgroup analysis, using the mixed model (method of moments) for levels of significance. A P value of less than 0.05 was considered to be significant. Sensitivity analysis was performed in each subgroup by running the analysis several times and each time removing one study. No outliers could be identified in this analysis.

For data pertaining to training case selection, two authors extracted conversion rates for different risk factors (age, BMI, sex, neoplasia, ASA score, presence of intra-abdominal fistula, and previous open abdominal surgery) from multiple expert series and computed odds ratios using Review Manager [Version 4.2.10 (November 13, 2006)]. Assuming heterogeneity among these studies, we applied the DerSimonian-Laird random-effects method for dichotomous data. Heterogeneity was tested by integrating Cochrane’s Q into the I²-formula. Nonpublished data for certain studies were obtained directly from the investigators.

Quality Assessment

It was not possible to apply a classical bias risk assessment method for the included articles as no randomized controlled trials were identified. Two reviewers independently assessed these studies by applying the validated Newcastle-Ottawa scale for cohort studies for bias risk assessment. For a more comprehensive quality assessment, we used the “signal-to-noise” concept. The noise represents the methodological weakness of a study (1-method score). The signal represents the study in the context of the review question and contains 3 key elements: relevance, applicability, and effect size. For each of these 3 components, there were 3 further items to rate (Supplemental Digital Content 2). By multiplication of the relatives of signal and method (1-noise) score, we computed a “signal-to-noise ratio” with a score range of 0 to 1. Thus, only articles with maximum signal and minimum noise would score 1, whereas studies with either minimal signal or maximal noise would score 0, independent of the size of the other factor. Interrater reliability was calculated using Bland-Altman analysis. Because this scoring system is yet to be validated, it was used merely for descriptive purposes, and, therefore, studies were included or excluded independent of their signal-to-noise ratio. However, none of the included studies scored 0.

RESULTS

After removing duplicates, we retrieved 3420 citations from the literature search (MEDLINE 3060 studies, EMBASE 360 studies, PsycINFO no studies). One hundred fifteen reports potentially contained 3 key elements: relevance, applicability, and effect size. Each of these 3 components, there were 3 further items to rate (Supplemental Digital Content 2). By multiplication of the relatives of signal and method (1-noise) score, we computed a “signal-to-noise ratio” with a score range of 0 to 1. Thus, only articles with maximum signal and minimum noise would score 1, whereas studies with either minimal signal or maximal noise would score 0, independent of the size of the other factor. Interrater reliability was calculated using Bland-Altman analysis. Because this scoring system is yet to be validated, it was used merely for descriptive purposes, and, therefore, studies were included or excluded independent of their signal-to-noise ratio. However, none of the included studies scored 0.

TABLE 1. Comparison of Pooled Weighted Outcomes of Mentored and Expert Series

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mentored*</th>
<th>Experts*</th>
<th>P-value 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversions</td>
<td>0.13 (0.10–0.17)</td>
<td>0.10 (0.77–0.13)</td>
<td>0.2853</td>
</tr>
<tr>
<td>Complications</td>
<td>0.20 (0.16–0.24)</td>
<td>0.25 (0.19–0.31)</td>
<td>0.4933</td>
</tr>
<tr>
<td>Anastomotic leak</td>
<td>0.03 (0.01–0.06)</td>
<td>0.04 (0.02-0.08)</td>
<td>0.3682</td>
</tr>
<tr>
<td>Mortality</td>
<td>0.02 (0.01–0.04)</td>
<td>0.02 (0.01–0.02)</td>
<td>0.5680</td>
</tr>
</tbody>
</table>

*Pooled mean rates, random effects model (95% confidence intervals).
†Mixed effect model (method of moments).

We excluded a total of 79 reports. Forty-nine studies had nonusable or nonrelevant data. In 28 studies, there was insufficient data on initial experience (“early learning curve”). One study on laparoscopic rectum resections was excluded as the surgeons performing 36 of the first 50 resections had already previously performed more than 50 colon resections. Further study could not be included as the investigators were unable to provide more detailed data on individual patients. A table of excluded studies may be requested from the authors; a table with the full description of all included studies is available on the Internet (Supplemental Digital Content 3).

The average signal-to-noise ratio was 0.43 (0.22–0.70), the method score 0.65 (0.44–0.83), and the signal score 0.66 (0.46–0.88). The interrater reliability showed a variation of only +0.01 (−0.22 to 0.25) in the Bland-Altman analysis (Fig. 2).

Mentored Versus Nonmentored Training

Five studies addressed the impact of a mentor’s physical presence in the operating department on the trainees’ performance, taking mortality, and morbidity data as end points from a total of 713 cases. A further 2 studies, with 38 cases, assessed the effect of a mentor present in the operating department only at the beginning of the training phase, and then with subsequent training via a remote video-link device for the further cases (telementoring). No publication bias was detected for these studies as no significant asymmetry was shown on the funnel plot. A comparison of the pooled weighted outcomes of all mentored cases (n = 751) compared with those of large expert case series (n = 5313) showed no significant difference in conversion, anastomotic leak, or mortality rate (Table 1). Studies on mentored programs were compared with the early case series (n = 695) of nonmentored (self-taught) surgeons in a similar way. Conversion rates of mentored trainees and expert surgeons (13% vs 11%, mixed model P = 0.005) and, therefore, studies were included or excluded independent of their signal-to-noise ratio. However, none of the included studies scored 0.

Selection of Training Case

We identified 8 studies that focused on training case selection. One was a survey study among expert surgeons that aimed to rank different types of resection in terms of difficulty, and overall rated right-sided colonic resections as easier than left-sided. Another investigated the effect an increased visceral to body fat index, rather than simply a high BMI, had on the rate of conversion. Six studies on case selection with a total of 4444 cases were used in a meta-analysis to delineate predictive risk factors for conversion and thereby determine case difficulty. Studies on mentor’s physical presence in the operating department on the trainees’ performance, taking mortality, and morbidity data as end points from a total of 713 cases. A further 2 studies, with 38 cases, assessed the effect of a mentor present in the operating department only at the beginning of the training phase, and then with subsequent training via a remote video-link device for the further cases (telementoring). No publication bias was detected for these studies as no significant asymmetry was shown on the funnel plot. A comparison of the pooled weighted outcomes of all mentored cases (n = 751) compared with those of large expert case series (n = 5313) showed no significant difference in conversion, anastomotic leak, or mortality rate (Table 1). Studies on mentored programs were compared with the early case series (n = 695) of nonmentored (self-taught) surgeons in a similar way. Conversion rates of mentored trainees and expert surgeons (13% vs 11%, mixed model P = 0.2835, fixed model P = 0.0127) (Table 2 and Fig. 3).

Simulation

There were several studies on reconstructed nonlive animal models, simulation with synthetic tissue, computer-based learning,
FIGURE 2. Equation for the signal-to-noise ratio and interrater reliability (Bland-Altman plot).

**TABLE 2.** Comparison of Pooled Weighted Conversion Rates for Mentored, Nonmentored and Expert (Control) Series [mixed effect model (Method of Moments)]

<table>
<thead>
<tr>
<th>Statistical Model</th>
<th>Mentored (n = 751)</th>
<th>Expert (n = 5313)</th>
<th>Nonmentored (n = 665)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects</td>
<td>0.14 (0.12–0.17)</td>
<td>0.12 (0.11–0.13)</td>
<td>0.22 (0.19–0.25)</td>
</tr>
<tr>
<td>Random effects</td>
<td>0.13 (0.10–0.17)</td>
<td>0.12 (0.10–0.13)</td>
<td>0.21 (0.15–0.26)</td>
</tr>
<tr>
<td>P</td>
<td>0.2835</td>
<td>0.0332</td>
<td>0.0002021</td>
</tr>
</tbody>
</table>

and cadaveric and animal LCS training courses. Of these, only 4 had extractable data, which pertained to candidate satisfaction, and these were included for analysis. There were no reliable data obtained for fidelity (tissue property, anatomical landmarks, environmental context). And again, unlike for basic laparoscopic skills in other specialties, objective data on the impact of the course on clinical skills were not demonstrated. Having adjusted the satisfaction scores to comparable ranges of Likert scale, the synthesis of data showed that participants of courses with cadaveric models were highly satisfied with the teaching value and the reliability of the materials used [mean 4.5 (range, 4–5)]. Trainees on the animal (pig) course, by comparison, reported considerably less favorable feedback [2.15 (range, 1.5–3)]. The number of delegates performing LCS increased from 53% to 81% after attending a cadaveric training course, with 26% of graduates performing, on average, one colectomy a week. One study demonstrated construct validity of a synthetic tissue-based simulator with expert surgeons performing significantly better than novices, and another showed a self-perceived increase in resident’s knowledge after completing a multimedia course.

**Assessment**

There were 3 studies on the implementation of assessment tools in LCS training. A global assessment score, which has been partially validated previously, was used to evaluate the trainee’s ability in 2 studies, and the completion of the operation in another. The studies using a global assessment score show good interitem and interrater reliability [Cronbach $\alpha = 0.88$ and 0.82, respectively, interrater reliability 0.76 (Pearson correlation)]. None of the studies tested specifically for predictive validity, acceptability, educational impact, or cost-effectiveness.

**DISCUSSION**

This review shows that mentored trainees achieve good clinical results in LCS. Surgeons perceive that teaching on cadaveric models provides a realistic method for training in a simulated environment. There remains a lack of data on the assessment of technical skills.

Our study is the first report with a comparative meta-analysis of clinical outcome data in training and was possible only by retrieving nonpublished data directly from the investigators. Analyzed data of 6064 patients reflect that trainees with an appropriate level of supervision generate the same complication, conversion, and mortality rates as expert surgeons. Similar results have been shown for carotid endarterectomy in a large, single-center study.

It is difficult to measure accurately manual performance during training. In this review, only 3 studies fell into the category of...
FIGURE 3. (a and b) Summary of conversion rates and 95% confidence intervals for nonmentored and mentored cases (fixed and random effects model). By comparison, the expert group that included a total of 5313 cases had a conversion rate of 0.12 (95% CI: 0.10-0.13) and was not different to the mentored case group (see also Table 2).
assessment of technical skills in LCS. Two demonstrated the use of dedicated global assessment tools during the training period of laparoscopic colorectal surgeons. Such scales are practical and easy-to-use and can help to estimate the performance level of a trainee. However, despite being partially validated and used widely, these scales have limitations as they demonstrate a ceiling effect and lack precision. Such tools often focus on basic surgical skills, which should already have been mastered by an advanced trainee. These issues may be addressed by human reliability methods that analyze errors enacted during procedures and identify the underlying performance shaping factors. Nevertheless, these methods remain in their infancy. Another approach is morbidity and mortality data, which give a global indicator of performance, but here again lie several limitations. First, audit results can be assessed only when both large numbers and long-term follow-up data are available. This takes time and does not avoid unfavorable outcomes accumulating while the audit is in progress. Second, surgical competence is multidimensional and includes operative skills, cognitive factors, personality traits, and decision making. Third, morbidity and mortality data cannot provide a prescriptive method for specifying how the performance of a procedure can be improved. Despite these shortfalls, an audit process remains essential, as it promotes monitoring to ensure the maintenance of safe clinical practice throughout training. In many studies, conversion rate was used as a performance parameter, although some authors state that this should not be considered to be a complication in itself, because the decision to convert may also reflect an experienced surgeon’s sound judgment. Nevertheless, it is expected for surgeons to have similar complication and conversion rates at the same stage of training. Interestingly, this review demonstrated a significant difference in conversion rates between mentored and nonmentored trainees and hence this is likely to represent a difference in the technical ability to complete a case laparoscopically. Although trainees at this level are often already proficient in basic technical skills, an experienced trainer may further aid intraoperative decision making and the comprehension of anatomy, and guide the trainee to utilize error reduction mechanisms and thus potentially minimize the rate of unnecessary conversion. Patient factors leading to conversion were analyzed and taken as indicators of case difficulty, thus allowing for appropriate training case selection in the early learning curve. Furthermore, for the first time, the meta-analyzed data provide enough supportive evidence to show that previous abdominal surgery is also a significant risk factor for conversion.

Our results reflect the effects of cognitive apprenticeship and the scaffolding instruction strategy on performance. Cognitive apprenticeship is the combined process of simultaneously teaching craft and strategic thinking. It describes the process of teaching in the operating room more accurately than mentoring, a term that has been used by most authors of the aforementioned studies. Scaffolding as a teaching strategy is a concept that is unconsciously or consciously used by most surgical teachers as they provide individualized support according to the trainee’s abilities. This can also be reflected by an expert’s careful choice of a suitable training case appropriate to the trainee’s level of skill. The meta-analysis of data from 3430 cases for the risk of conversion, representing a predictive level of difficulty, may aid this selection process. Vygotsky described the Zone of Proximal Development as the range of tasks someone is able to learn. If the teacher deploys the scaffolding instruction strategy, the trainee is able to move up along this scale, and in this context our results can be explained.

Novel simulation techniques designed specifically for LCS have been developed. These include a combination of virtual reality simulators and box trainers, animal and human tissue, and synthetic materials. This systematic review revealed a notable lack of available data on the educational value of simulated training in LCS. Unlike for laparoscopic cholecystectomy and basic gynecological procedures, the transferability to clinical training has not been investigated scientifically for these methods. The only identifiable studies on LCS courses were satisfaction surveys that indicated that cadaveric models seem to be superior, in terms of reality and learning effect, to live animal (pig) models, despite the lack of tissue perfusion. The recent studies on hybrid simulation and an initial report on virtual reality colonic models based on individual patient’s imaging data may advance simulation training as they will not only negate the availability and ethical issues surrounding cadavers but also allow a stepwise increase in task difficulty similar to clinical training.

This review also has its limitations. First, a direct comparison between the 3 groups (mentored, nonmentored, and expert) may be problematic as there was no case control for surgical complexity and patient morbidity. Expert surgeons are more likely to operate on more complex cases that may explain why no significant difference between mentored trainees and experts was found. Nevertheless, a significant difference was found on comparing the mentored and nonmentored subgroups. Second, difficulties exist when synthesizing nonmentored subgroups. Second, difficulties exist when synthesizing educational data on simulated training and clinical outcome. In particular, no studies compared the effects of different simulation methods on training. Also, there is no available study comparing the cadaveric and animal models by the same trainees that would provide more direct and reliable information.

The collated evidence provided by this review permits the suggestion of several considerations when setting up a training program for technically advanced procedures. First, supervised training and appropriate case selection even at an advanced level is essential to ensure safe clinical practice. Further studies need to evaluate the training effect on the proficiency gain process. Second, integrated simulation models are more likely to be effective when anatomically accurate and realistic. From the trainee’s perspective, cadavers fulfill these requirements more than animal models. Third, a training program needs to be evaluated and have a governance structure. This is particularly important in the context of the modern world where patient safety is paramount. It is also relevant in terms of medicolegal indemnity issues, as a higher complication or conversion rate as the result of a learning process is no longer acceptable. Ideally,

### TABLE 3. Meta-analysis of Odds Ratios for Risk Factors for Conversion (Random Effects Model)

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>N*</th>
<th>n*</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &gt; 64 yr</td>
<td>2444</td>
<td>2340</td>
<td>1.15 (0.68–1.94)</td>
<td>0.61</td>
</tr>
<tr>
<td>Male sex</td>
<td>4444</td>
<td>2108</td>
<td>1.72 (1.40–2.11)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>PAS</td>
<td>2219</td>
<td>659</td>
<td>1.65 (1.19–2.98)</td>
<td>0.003</td>
</tr>
<tr>
<td>BMI &gt; 30</td>
<td>3684</td>
<td>513</td>
<td>1.66 (1.26–2.20)</td>
<td>0.0003</td>
</tr>
<tr>
<td>ASA 3 and 4</td>
<td>2684</td>
<td>614</td>
<td>1.737 (1.331–2.266)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cancer</td>
<td>3430</td>
<td>1047</td>
<td>1.798 (1.443–2.241)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Fistula</td>
<td>2482</td>
<td>87</td>
<td>4.229 (2.575–6.944)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* N, total patients; n, group exposed to risk factor.

ASA indicates American Society of Anesthesiologist (score); BMI, body mass index (BMI) (kg/m²); 95% CI, 95% confidence interval; OR, odds ratio; PAS, previous abdominal surgery.
such an evaluation should encompass both clinical outcome data to ensure quality and trainee performance data to monitor the training progress. Further research for advanced intervention and surgical training programs needs to focus on the educational value of different simulation methods and the predictive value of assessment tools. Although trainees seem to appreciate the benefits of cadaveric training, the availability of such models is limited. Novel, computer-based simulation needs to be assessed for training effectiveness. The ideal feasible assessment tool is still to be developed.

ACKNOWLEDGMENTS

We thank Daniel Birch (Edmonton, Alberta, Canada), Selim Dincler and Peter Buchmann (Zurich, Switzerland), Omar Faiz (London, United Kingdom), In Ja Park (Daegu, South Korea), Nicolas Rotholtz (Buenos Aires, Argentina), and Christopher Schlachta (Toronto, Ontario, Canada) for providing us with unpublished data of their series. We also thank Elena Kalinskaya, Director of Statistical Advisory Service, Imperial College, London, for her expert advice and review of the statistical methods. We are grateful to Matthias Egger, Head of the Department of Epidemiology and Public Health University of Bern (Switzerland), for his helpful comments on an earlier draft of this manuscript.

D.M., S.M.W., and G.B.H. designed the study and drafted the manuscript. D.M., S.M.W., and M.N. acquired and analyzed data. G.B.H. and A.W.D. critically revised the manuscript.

REFERENCES


