Low Quality of Evidence for Robot-Assisted Laparoscopic Prostatectomy: Results of a Systematic Review of the Published Literature

Diana C. Kang\textsuperscript{a}, Miranda J. Hardee\textsuperscript{b}, Susan F. Fesperman\textsuperscript{a}, Taryn L. Stoffs\textsuperscript{a}, Philipp Dahm\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a}Department of Urology, University of Florida, Gainesville, FL, USA
\textsuperscript{b}Department of Surgery, Division of Urology, University of Utah, Salt Lake City, UT, USA

\textbf{Abstract}

\textbf{Background:} Robot-assisted laparoscopic prostatectomy (RALP) is displacing radical retropubic prostatectomy as the gold standard surgical approach for clinically localised prostate cancer in the United States and is also being increasingly used in Europe and other parts of the world. This trend has occurred despite the paucity of high-quality evidence to support its relative superiority to more established treatment modalities.

\textbf{Objective:} We performed this study to critically assess the quality of published evidence on RALP to support this major shift in practice patterns.

\textbf{Design, setting, and participants:} We conducted a systematic review of the published literature through Medline and Embase (1966 to December 2008). All original research publications on RALP were included. Editorials, letters to the editor, and review articles were excluded.

\textbf{Measurements:} Two reviewers independently performed the data abstraction using a standardised form derived from the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) criteria.

\textbf{Results and limitations:} Seventy-five original research publications met eligibility criteria. Fifty-five (73.3\%) studies were published between 2005 and 2008, and 20 studies (26.7\%) were published between 2001 and 2004. Approximately three-quarters of the studies were case series (74.7\%), and only two (2.7\%) randomised, controlled trials (RCT) were identified. Twelve authors cowrote 72\% (54 of 75) of the published studies. Reporting of STROBE criteria ranged from 100.0\% (scientific rationale/background explained) to 1.3\% (consideration of sample size), with no improvement over time. The study was limited to published literature in the English language.

\textbf{Conclusions:} The published RALP literature is limited to observational studies of mostly low methodologic quality. Our findings draw into question to what extent valid conclusions about the relative superiority or equivalence of RALP to other surgical approaches can be drawn and whether published outcomes can be generalised to the broader community. There is an urgent need to raise the methodologic standards for clinical research on new urologic procedures and devices.

\textsuperscript{*}Corresponding author. Department of Urology, University of Florida, College of Medicine, Health Science Center, Box 100247, Room M2-204, Gainesville, FL 32610-0247, USA. Tel. +352 273 6815; Fax: +352 273 8335. E-mail address: p.dahm@urology.ufl.edu (P. Dahm).
1. Introduction

The first robotic-assisted laparoscopic prostatectomy (RALP) was performed in Frankfurt, Germany, in 2000 by Binder et al using the da Vinci Surgical System (Intuitive Surgical, Sunnyvale, CA, USA) [1]. Since that time, RALP surgery has made a remarkable impact on how radical prostatectomies (RP) are performed in the United States as well as in other industrialised countries. According to the manufacturer, approximately 70% of all prostatectomies in the United States in 2007 were performed as RALP procedures, with every indication that this number will continue to increase [2]. Other countries are observing similar trends. Urologists have become frequent users of the da Vinci Surgical System, with RPs becoming the most common robotically assisted surgical procedure today [3]. Claims of superior outcomes compared to other surgical approaches include lower blood loss, decreased length of hospital stay, shorter convalescence, as well as better cancer control as measured by lower rates of positive surgical margins and improved functional outcomes [4,5].

However, it has been noted in a recent and well-done systematic review by Ficarra et al that these claims of superior clinical outcomes are largely based on retrospective observational studies that are not well suited to provide high-quality evidence to address therapeutic effectiveness [6,7]. This conclusion casts doubt upon the true advantages of RALP compared to other surgical approaches [7,8]. Two population-based studies by Hu et al have drawn into question the relative superiority of RALP with regards to cancer control, urinary continence, and erectile function [9,10]. Questions regarding the comparative effectiveness of RALP appear particularly relevant in light of the high costs associated with the installation and maintenance of a robotic surgical system [11]. The purpose of this study was therefore to examine the quality of evidence underlying the paradigm shift from open to robot-assisted prostatectomy while at the same time serving as a case study for the evidence standards for new technology being adopted by the urologic community.

2. Materials and methods

A systematic literature search was initially conducted in August 2007 and updated in November 2009 using both the Medline and Embase databases to identify all original research publications relating to the use of RALP in human subjects published from 1966 to December 2008 (inclusive). The following search terms were used: radical prostatectomy, da Vinci, daVinci, robot, robotic, and RALP. Search limitations included human-based studies and publication in the English language. Studies identified were screened for eligibility by two independent reviewers. Exclusion criteria included editorials, letters to the editor, and review articles. Systematic reviews were used only if the study included original research data from the publishing institution. A final consensus on study selection was achieved by discussion between the two reviewers and arbitration by a third reviewer.

A standardised data-abstraction form that included surgery-relevant criteria from instruments previously reported in the published literature [12], an extension of the Consolidated Standards of Reporting Trials (CONSORT) statement to randomised trials on nonpharmacologic treatments [13], and all criteria of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [14] were used. Both the CONSORT and the STROBE statements represent established standards for transparent and high-quality reporting of important items relating to the methodology, analysis, result reporting, and discussion of randomised trials and observational studies (cohort, case-control, and cross-sectional studies), respectively. A total of 48 criteria were included regarding general study characteristics, institutional review board or ethics committee study approval, description of surgeons and institutions, background and intervention, statistical methods, and description of results and outcomes. Studies were furthermore rated for the level of evidence provided according to criteria by the Centre for Evidence Based Medicine in Oxford, United Kingdom [15]. Prior to review, the data-abstraction form was pilot tested in a separate sample of six articles from the robot-assisted laparoscopic cystectomy literature to ensure the greatest possible interobserver agreement. Data abstraction was then performed on the 75 publications using the standardised data-abstraction form and entered into a dedicated database. Interobserver agreement was recorded, and discrepancies were resolved through discussion and arbitration by a third investigator to generate a consensus version.

Statistical analysis was performed using SPSS v.17.0 (SPSS, Chicago, IL, USA). Interobserver agreement beyond chance was assessed for a subset of 12 predefined variables using the $\kappa$ statistic. We tested the a priori hypothesis that the methodologic and reporting quality of studies improved from 2005–2008 compared to the initial period of 2001–2004, representing two time periods of equal length (4 yr). In an additional post hoc comparison, we limited the same analysis to the subset of studies that were written by investigators who contributed to more than one study. All statistical testing was performed two-sided with a predefined $\alpha$ of 0.05. No formal adjustment for multiple comparisons was used.

3. Results

A systematic literature search identified 213 published articles eligible for review. Of these, 75 studies met inclusion criteria and were included in this analysis (see Appendix), the first of which was published in 2001. Table 1 summarises the baseline characteristics of these studies. Fifty-five (73.3%) studies were published between 2005 and 2008, and 20 studies (26.7%) were published between 2001 and 2004. Approximately three-quarters were case series (74.7%), and only two (2.7%) randomised, controlled trials (RCT) were identified. These two RCTs compared technical modifications of RALP with each other rather than comparing RALP to open approaches. Overall, the published literature provided level 2, 3, and 4 evidence in 2.7% ($n = 2$), 24.0% ($n = 18$), and 73.3% ($n = 55$) of studies, respectively. The median sample size was 120 patients (interquartile range: 55–300), with only 12.0% of studies reporting on >500 patients and only three studies (3.9%) including ≥1000 patients. The vast majority of studies originated from a single centre (96.0%) and did not report their source of funding (97.3%). Twelve authors cowrote 72% (54 of 75) of the published studies. Only 4 of 10 studies explicitly reported study approval by their local ethics committee or institutional review board.

Tables 2–4 summarise the reporting of study methods, results, and discussion of these studies. The mean $\kappa$ as a measure of interobserver agreement beyond chance was 0.81, with a range of 0.54 (learning curve addressed) to 1.00 (scientific rationale/background explained). A high proportion of studies provided the scientific rationale for the study...
that described their surgical technique dropped from 100.0% to 72.2% \((p = 0.009)\), and addressing the learning curve fell from 90.0% to 54.2% \((p = 0.005)\).

Table 3 provides summary information on the quality of result reporting. A large proportion of studies described the study population overall \((70.7\%)\), stated the timeline for when the procedures were performed \((76.0\%)\), and described the number and nature of complications \((76.0\%)\). Less than half of the studies \((41.2\%)\) provided an average length of follow-up, and less than a quarter of the studies \((22.7\%)\) addressed completeness of follow-up. Reporting did not improve over time for any of these criteria.

Information about reporting quality of discussion items is summarised in Table 4. Strong points were the interpretation of study results \((98.7\%)\), addressing the underlying hypothesis \((90.7\%)\), and interpretation in the context of the current evidence \((97.3\%)\). Meanwhile, less than one-quarter of the studies addressed generalisability of the results \((21.3\%)\), approximately one-third discussed the implications of missing data \((36.1\%)\), and only half included a discussion of potential sources of bias in the study \((50.7\%)\). With regard to the quality of reporting of discussion criteria, no significant changes over time were noted.

4. Discussion

In this study, we critically appraised the quality of evidence underpinning the dramatic paradigm shift from open RP to RALP as the preferred surgical approach. Key findings of this study were the absence of any RCT that compared the therapeutic effectiveness and potential complications of RALP to the more established surgical approaches as well as the lack of improvement of reporting quality over time. The only two RCTs we found contrasted technical modifications of RALP rather than comparing RALP to open or laparoscopic prostatectomy. Both studies, which focused on the effectiveness of the so-called Rocco stitch and the relative advantages and disadvantages of an extra- versus intraperitoneal approach were reported in the 2005–2008 time period by major contributors to the RALP literature \([16,17]\). However,
the results of our analysis did not differ (data not shown) when limited to the subset of 54 studies by authors who had published more than one study. Our findings are largely consistent with those of a recent systematic review by Ficarra et al of the open retroperitoneal, laparoscopic, and RALP procedures. The authors found that laparoscopic prostatectomy and RALP were associated with lower blood loss and transfusion rates but concluded that there was little evidence to support relative superiority of one operative approach over another overall [7].

RCTs allow investigators to carefully control for different types of bias by implementing methodologic safeguards such as random allocation concealment, blinding, intention-to-treat analysis, and completeness of follow-up [18]. For these reasons, RCTs (and systematic reviews thereof) stand at the top of the hierarchy of evidence, which represents one of the guiding principles of evidence-based clinical practice [19]. Within the framework of an evidence-based practice of urology, one would therefore expect that the widespread and rapid adoption of a new surgical device be supported by high-quality evidence, indicating superiority, which was not the case for RALP. The way new technology is adopted—as exemplified by RALP—therefore stands in notable contrast to the standards for pharmaceutical agents, where RCT evidence of at least noninferiority is a prerequisite for regulatory approval—for example, by the US Food and Drug Administration (FDA). In contrast, standards for FDA approval of surgical devices are much less rigorous, thereby underscoring the role of aggressive marketing to promote the adaptation of new technology in the community.

RCTs of surgical procedures and devices can be particularly challenging. Specific issues relate to the inability to blind surgeons to the procedure being performed as well as the logistical and ethical challenges of blinding study subjects [18]. Another issue relates to the difficulty in controlling for the surgical skills and experience of individuals surgeons, which have been shown to be an important predictor of outcomes [20, 21]. In addition, surgeons and patients frequently lack equipoise and are therefore unwilling to accept random allocation to an invasive surgical procedure [12]. These reasons in part explain the relative paucity of RCTs in the urologic literature, thereby emphasising the importance of observational studies of high methodologic quality [12, 22]. With this background, the low reporting quality of observational studies, which represents the vast majority of evidence for RALP, is critically important. Not only were a majority of studies single-armed surgical case series that lacked a comparison group, but they also failed to provide the type of information that would allow critical readers to assess the validity, impact, and applicability of the study results to their patients. For example, failure of study authors to address the handling of missing data as well as the lack of information on completeness of follow-up raise concerns about selection bias. Also, failure to identify study inclusion and exclusion criteria as well as origin of nearly three-quarters of studies from only 12 centres question to what extent the results can be generalised to a larger population.

These concerns are underscored by a recent population-based study by Hu et al, which found that patients undergoing minimally invasive prostatectomy (laparoscopic and robotic-assisted laparoscopic) compared to open techniques had 3.7 times (95% confidence interval, 2.8–4.8) increased odds of requiring salvage therapy [23]. A follow-up study also found an increased risk of complications, urinary incontinence, and erectile dysfunction in patients undergoing a form of minimally invasive prostatectomy from 2003 to 2007 [9]. Other issues with RALP are the extreme costs related to the initial installation and maintenance of a robotic system [11]. These findings raise questions as to whether the rapid and relatively uncritical adoption of RALP was appropriate when viewed from a societal perspective. The
importance of this issue is further underscored by the fact that comparative effectiveness research of robot-assisted surgery was identified as one of 100 priority topics identified by the American Recovery and Reinvestment Act of 2009 [24].

Strengths and limitations of this study deserve consideration. Strengths include the use of a data-abstraction instrument that was adapted from the STROBE checklist, pilot tested and independently applied by two separate reviewers who achieved a high degree of interobserver agreement. Limitations include our focus on studies published in the English language. Second, we assumed that only studies that resulted in full-text publication would provide the necessary methodologic detail to assess whether their claims were valid. Third, we acknowledge the inherent subjectivity of any assessment of methodologic quality, even when performed by two independent observers. In this study, we applied minimal standards for each of the STROBE criteria, giving investigators the benefit of the doubt. Our results may therefore be regarded as an optimistic interpretation of the actual reporting quality. Finally, we recognise that the actual methodologic study quality may be better than reported [25]. However, because the full-text publication represents the only readily accessible record of how a study was conducted, assessments such as these remain relevant.

In summary, this study finds that rapid and widespread adoption of RALP is not supported by high-quality evidence that would be suitable to demonstrate relative superiority over alternative surgical techniques it has replaced. In an era of exploding health care costs that mandate an increasing role for comparative effectiveness research, higher standards for the adoption of new surgical devices appear necessary. Aside from stricter regulatory requirements, raising the awareness for the principles of valid clinical research in the urologic community appears important. A recent article by McCullough et al that makes recommendations for the future assessment of surgical procedures and devices based on a five-stage description of the surgical development process is noteworthy in this context [26]. Ongoing efforts to conduct RCTs, such as the Randomised Controlled Trial of Laparoscopic, Open and Robot Assisted Prostatectomy as Treatment for Organ-confined Prostate Cancer (LopeRA) trial and the Bladder cancer: Open versus Laparoscopic or Robotic cystectomy (BOLER0) trial—both in planning in the United Kingdom—also lend hope that the role of robotic surgery will indeed be better defined in the future.

5. Conclusions

The published RALP literature is limited to uncontrolled, observational studies, many of which are of poor methodologic quality. Most studies originate from a few select centres. These findings draw into question to what extent valid conclusions about the relative superiority or equivalence of RALP to other surgical approaches can be drawn and whether outcomes can be generalised to a broader community. There is an urgent need to raise the methodologic standards for clinical research for newly introduced urologic procedures and devices.

Appendix

The original research studies (n = 76) listed below were included in the systematic review. One study was published in two parts (references 30 and 31) and was considered a single study for the purpose of this analysis.


Author contributions: Philipp Dahm had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Dahm.
Acquisition of data: Kang, Hardee.
Analysis and interpretation of data: Stoffs, Fesperman, Dahm.
Drafting of the manuscript: Kang.
Critical revision of the manuscript for important intellectual content: Kang, Dahm.
Statistical analysis: Stoffs, Dahm.
Obtaining funding: None.
Administrative, technical, or material support: Stoffs.
Supervision: Dahm.
Other (specify): None.

Financial disclosures: I certify that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: None.

References