Methodological Review by David C. Chang, Ph.D., MPH, MBA

Article Reviewed

Hu J et al. Comparative Effectiveness of Minimally Invasive vs Open Radical Prostatectomy. JAMA 2009; 302:1557-64

Review

The use of minimally invasive radical prostatectomy (MIRP) has surged in this decade, driven in part by patient perception which, in part, was driven by direct-to-consumer advertising. However, comparative effectiveness research between MIRP and open retropubic radical prostatectomy (RRP) has yet to be properly conducted. In the absence of randomized controlled trials, the authors performed a population-based observational study using the SEER-Medicare database.

Are the results valid?

This is a retrospective analysis of prospectively collected data, and is thus a prospective observational study and can be classified as a Level II study. The source of data is the US Surveillance, Epidemiology, and End Results (SEER) / Medicare-linked data¹ from 2003 through 2007. Inclusion criteria were men with prostate cancer who underwent MIRP vs RRP. Main outcome measures were 30-day complications, anastomotic stricture 31 to 365 days postoperatively, long-term incontinence and erectile dysfunction more than 18 months postoperatively, and postoperative use of additional cancer therapies (as a surrogate for cancer control).

The primary difference between a Level I study (randomized controlled trial) and a Level II study (such as this prospective observational study) is how potential confounders were controlled for in the study. Whereas a Level I study controls for confounders via randomization, a Level II study controls for confounders via multivariable regression analyses, by statistically adjusting for confounding variables.² This allows for the comparison of different patient populations, by mathematically accounting for the effects of different variables on outcomes. And so the validity of a Level II study rests primarily on how comprehensive this adjustment is. In this study, covariates accounted for included: age; race/ethnicity; marital status; patient comorbidity using the Klabunde modification of the Charlson index;³ preoperative diagnoses of incontinence and erectile dysfunction; surgeon volume; census tract measures of median household income and proportion of individuals with at least a high school education; SEER region; population density (rural vs urban); and calendar year of the surgery. This is a comprehensive list, including potential confounders at the patient level, surgeon level, and regional level.⁴

In addition to controlling for confounders of outcomes, this study also controls for confounders that may influence the likelihood of different patients to undergo MIRP vs RRP (the so-called “treatment selection bias”) via the propensity score method.⁵ In essence, the propensity score mathematically creates a pseudo-randomization process to attempt to balance patient characteristics between groups.

Another important determinant of validity is the patient selection, or whether the selected patients are representative and generalizable. Here, a population-based study such as this one has an advantage over a randomized controlled trial: it includes more patients, and more different types of patients, than any randomized controlled trial could possibly include.
In summary, the study has high validity because it controls for potential confounders on outcomes through multivariable regression analysis; it controls for differential likelihood of patients to receive MIRP vs RRP via propensity score method; and it selects for a large and generalizable patient population.

*What are the results?*

MIRP patients experienced the following improved outcomes in comparison to RRP: shorter length of stay, fewer respiratory complications, fewer miscellaneous surgical complications, fewer strictures. However, MIRP patients experienced the following worse outcomes: more genitourinary complications, incontinence, and erectile dysfunction. There was no difference in the use of additional cancer therapies.

*Can the results be applied to patient care?*

This study provides valuable data to assist patients in deciding whether to choose MIRP vs RRP. The widespread direct-to-consumer advertising of MIRP may have subtly promoted publication bias against any study that may detail the challenges and suboptimal outcomes of MIRP, especially among surgeons early in the learning curve. This study provides a real-world and comprehensive look at the true outcomes of MIRP.

References

Clinical Review by Matthew R. Cooperberg, MD, MPH

Article Reviewed

Hu J et al. Comparative Effectiveness of Minimally Invasive vs Open Radical Prostatectomy. JAMA 2009; 302:1557-64

Review

As robot-assisted radical prostatectomy gains an ever-increasing share of prostatectomy procedures in the United States and worldwide, the efficacy of this procedure relative to open radical prostatectomy has become an increasingly important question. A large, recent systematic review found few differences among the approaches to prostatectomy in terms of oncologic or health-related quality of life (HRQOL) outcomes, concluding, perhaps charitably, that the quality of existing comparative studies is "not excellent." Indeed, a recent Institute of Medicine report on comparative effectiveness research identified open vs. robot-assisted prostatectomy as one of the top 100 national priority areas for future research.

The paper by Hu et al, published last year in JAMA, aimed to compare short-term complications, intermediate-term HRQOL outcomes, and oncologic outcomes across surgical modalities—using the administrative billing codes compiled in the linked Surveillance, Epidemiology, and End Results (SEER)-Medicare database. Because robot-assisted and non-robot-assisted laparoscopic prostatectomy are generally billed under the same CPT code, the two approaches cannot be distinguished in the database, and in the study were conflated as "minimally-invasive radical prostatectomy (MIRP)."

Are the results valid?

The study tackles a number of different questions, and the validity of the findings varies substantially with the question. In terms of capturing trends in uptake of MIRP, SEER-Medicare data are population-based, but by definition are restricted to men over 65 and are not fully representative of the U.S. geographically. The Detroit and California registries, for example, which account for two-thirds of the MIRP cases in the study, clearly reflect heavy influence of specific institutions which were early adopters and high volume practitioners of robot-assisted surgery during the study period. It is important to recognize that the study period represented the beginning of the learning curve for robot-assisted surgery for most of the surgeons whose cases were included. Particularly in this context of rapid technology migration, combining robot-assisted and non-robot-assisted laparoscopic cases clearly is not ideal.

Medicare data are an excellent source of information regarding length of stay and relatively short-term complications (cardiopulmonary events, anastomotic strictures, etc.) which generally will be reflected accurately in billing codes.

In terms of oncologic outcomes, SEER during the observation period included fairly marginal risk stratification data: AJCC pathologic stage and WHO grade, but no PSA data or Gleason scores. Oncologic failure was defined only by application of secondary radiation and/or hormonal therapy, timing of which varies substantially from practice to practice, and depends on various clinical risk factors. For locally advanced cases, including some with T2 disease but positive margins, radiation also may be given as an adjuvant in the absence of any evidence of recurrence. Finally, diagnosis dates were included through...
2005, but treatment was allowed through last day of the study observation period; in general the followup was quite short for most cases in the cohort.

The biggest methodological problems arise, however, in the paper’s attempt to use Medicare coding data to draw inferences regarding urinary and sexual HRQOL outcomes. Research published over a decade ago by Litwin et al demonstrated that physician reports inadequately capture HRQOL outcomes after prostate cancer treatment, and that these outcomes must be assessed via direct query to patients using validated questionnaires. A more recent study confirmed that this requirement is equally true in contemporary practice, and there really is very little controversy about this point in the prostate cancer HRQOL literature.

Numerous biases bear heavily coding practices for HRQOL domains, both at baseline and in followup. Fewer than half of new prostate cancer diagnoses are associated with documentation of baseline sexual function in the chart; the proportion with baseline erectile dysfunction whose billing forms include a secondary ICD-9 code to indicate this is presumably much lower. In followup, moreover, the decision to code “incontinence” or “impotence” may reflect only the loudness of a patient’s dissatisfaction with the outcome rather than the actual degree of impairment. Indeed, another recent study—incorporating well-validated, patient-reported HRQOL measures—found that undergoing robot-assisted rather than open prostatectomy was associated with substantially less satisfaction and greater regret—even with adjustment for actual HRQOL functional domains. Patients opting for robot-assisted surgery, particularly in the relatively early years of the technology’s promulgation included in the study by Hu et al, were quite likely for a variety of reasons to carry falsely optimistic expectations regarding their outcomes.

These reporting biases would not be corrected using propensity score-based modeling. Indeed, in the Hu et al study it is telling that there was no statistically significant difference by surgical modality in use of procedures for incontinence and impotence, which are coded far more consistently than HRQOL outcomes per se. While the authors briefly address the limitations of coding data in the discussion, their conclusions regarding the HRQOL outcomes of open prostatectomy vs. MIRP are stated far too strongly, particularly in the concluding statement of the abstract which drew the greatest attention from the lay press.

What are the results?

The authors confirmed a rapid increase—more than 4-fold—in use of MIRP over the study period from 2003 to 2007. In general MIRP patients were slightly less likely to be Black or Hispanic, and much more likely to be well-educated and high income earners.

Consistent with previous reports, MIRP was associated with shorter length of stay and lower rates of transfusion than open prostatectomy. MIRP was also associated with nearly 30% lower rates of respiratory complications and over 60% lower rates of anastomotic stricture, but higher rates of “other” genitourinary complications (the ICD-9 code list for these includes, for example, cystitis, pyelonephritis, and bladder/ureteral injuries, but there is no detailed breakdown provided for complications within the category). The perioperative mortality rate (0.16% across all patients) was low, and likely is an accurate reflection of contemporary practice among men over 65.

MIRP was associated with higher coding rates for incontinence (OR 1.3, 95% CI 1.1-1.6) and impotence (OR 1.4, 95% CI 1.1-1.7) compared to open prostatectomy, but no statistically significant difference in
procedures for either domain. Secondary treatment rates were not significantly different between MIRP and open surgery.

*Can the results be applied to patient care?*

The epidemiological observations on treatment trends are interesting and valid but have little bearing on clinical practice. Moreover, observed differences in short-term complications are unlikely to impact treatment decision-making. For the outcomes that matter most in prostatectomy—cancer cure and preservation of urinary and sexual function—the methodological shortcomings of this study obviate its relevance to clinical practice or health policy decision-making.

Robot-assisted radical prostatectomy should have been compared to open surgery in a randomized, controlled trial as interest was first rising in the procedure. Given the ongoing trends in practice patterns and surgical training, it is unfortunately now far too late for such a trial, at least in this country. However, the two surgical approaches can still be studied in carefully-designed, prospectively-accrued registries incorporating detailed clinical information and—critically—patient-reported HRQOL outcomes using validated instruments. It is only through collection of such high-quality data that the comparative effectiveness, and cost-effectiveness, of open vs. robot-assisted surgery can be fairly and accurately determined.

**References**