Evidence-based Urology

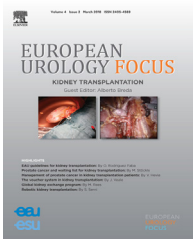
Puerto Nino, Angie K; van der Worp, Henk; Tikkinen, Kari A O; Guyatt, Gordon H; Blanker, Marco H

Published in:
European Urology Focus

DOI:
10.1016/j.euf.2021.09.024

Publication date:
2021

available at www.sciencedirect.com
journal homepage: www.europeanurology.com/eufocus



Mini Review – Epidemiology

Evidence-based Urology: Understanding Heterogeneity in Systematic Reviews

Angie K. Puerto Nino [a,*], *Henk van der Worp* [b], *Kari A.O. Tikkinen* [c,d], *Gordon H. Guyatt* [e,f], *Marco H. Blanker* [b]

[a] *Translational Research Program, Temerty Faculty of Medicine, University of Toronto, Toronto, Canada;* [b] *Department of General Practice and Elderly Care Medicine, University Medical Center Groningen, Groningen, The Netherlands;* [c] *Department of Urology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland;* [d] *Department of Surgery, South Karelian Central Hospital, Lappeenranta, Finland;* [e] *Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada;* [f] *Department of Medicine, McMaster University, Hamilton, Canada*

Article info

Article history:
Accepted September 29, 2021
Available online 26 October 2021
Associate Editor: Richard Lee

Keywords:
Systematic review
Heterogeneity
Evidence-based medicine

Abstract

Variability in the results of randomized trials presents challenges to the interpretation and application of the evidence to patient care. Understanding how systematic reviews deal with this problem of "heterogeneity" will help clinicians in applying results in their patient management. This manuscript offers a review of heterogeneity from the clinical urological perspective.

Patient summary: Systematic reviews of the literature are necessary to accurately summarize the available evidence to inform clinical decisions. In this mini-review, we explain how to understand and deal with the differences between studies—which we call heterogeneity—included in these types of reviews.

1. Introduction

Systematic reviews and meta-analyses—in urology as in other fields—critically appraise and synthesize the evidence on a specific health issue, facilitating the urologists' evidence-based decision-making [1]. Primary studies included in systematic reviews always differ to some extent in their design or results, and sometimes to a great deal.

Differences between studies are referred to as "heterogeneity". Heterogeneity is characterized as "clinical", referring to difference in participants' characteristics, interventions, and setting, or "methodological", including differences in risk of bias. "Statistical" heterogeneity refers to variability in the results of included studies—effects of intervention on a particular outcome may vary across studies from large to small, absent, or even harmful [2].

Urologists can identify clinical or methodological heterogeneity by comparing the included studies. Credible systematic reviews present tables describing the population characteristics, interventions, and risk of bias of each study. The best reviews anticipate these differences and develop a priori hypotheses (including the direction of effects) to explain differences in results. For instance, studies of prostate cancer treatment might hypothesize larger effects with more restricted disease, or higher risk of bias [3,4].

When considering statistical heterogeneity, we must first ask if chance can explain the degree of variability in study results. If the answer is yes, inconsistency does not threaten our confidence in the pooled estimates of effect. If the answer is no, we can ask if any of the authors' a priori hypotheses explain the heterogeneity.

* Corresponding author. Translational Research Program, Temerty Faculty of Medicine, University of Toronto, 263 McCaul Street, Toronto, Ontario M5T 1W7, Canada. Tel. +1 437 9296893.
E-mail address: apuertonino@gmail.com (A.K. Puerto Nino).

https://doi.org/10.1016/j.euf.2021.09.024

Forest plots—graphical displays of estimated results from several studies addressing the same question—can offer good visualization of statistical heterogeneity. The similarity of point estimates, the overlap in confidence intervals, and the relationship of the results to clinical decision thresholds—as we show later with examples from the literature—can inform the interpretation of results when determining heterogeneity [5].

Statistical methods for heterogeneity include the χ^2 test for heterogeneity and the I^2 statistic [1,2]. The smaller the *p* value associated with the test of heterogeneity, the less likely it is that chance can explain the results [2]: *p* values greater than 0.1 suggest that chance can easily explain differences between studies; values less than 0.1 but greater than 0.01 suggest that chance may not explain differences in results; and values less than 0.01 suggest that chance is a very unlikely explanation. I^2 ranges from 0% to 100%. I^2 values < 40% raise only minimal concerns; values of 30–60% raise more concerns; any result >50% raises appreciable concerns; and heterogeneity becomes very worrisome when I^2 exceeds 75% [2].

Authors can, and should, use subgroup analysis to see whether their a priori hypotheses regarding explanations of heterogeneity prove prescient. For instance, was—as hypothesized—the treatment effect really larger for patients with prostate cancer with more restricted disease? Did the studies at higher risk of bias actually demonstrate larger treatment effects? If the answer is possibly, authors may use criteria to decide on the extent of the credibility of the possible subgroup effects [2–4,6,7].

2. Evaluating heterogeneity

We illustrate the concepts in assessing heterogeneity using two recent meta-analyses. Many patients experience urinary incontinence after prostatectomy. The European Association of Urology 2020 incontinence guidelines provide a weak recommendation to use male slings in patients with mild to moderate postprostatectomy incontinence who have not responded to conservative management [8].

A 65-year-old male with one year of incontinence after prostatectomy presents at your clinic reporting persistent involuntary loss of urine on physical exertion and coughing (stress incontinence) after trying multiple conservative approaches. As a further intervention, he is interested only in an adjustable male sling and asks you how effective they are.

When searching the literature, you find one recent randomized trial that showed no major difference when comparing a male sling and an artificial urinary sphincter [9]. You also identify two systematic reviews on adjustable slings. Eighteen (95%) of the 19 studies in these reviews were single-arm studies reporting the success of a particular procedure—that is, there was only one very small study (22 patients) with a direct comparison. Thus, the risk of bias is high and any inferences will be limited. Nevertheless, this is the best evidence available, so you proceed.

The first review included 17 patient cohorts [10]. You search for differences in the patients enrolled in the studies and in the interventions applied. The study participants differed in the incontinence severity and the studies used five

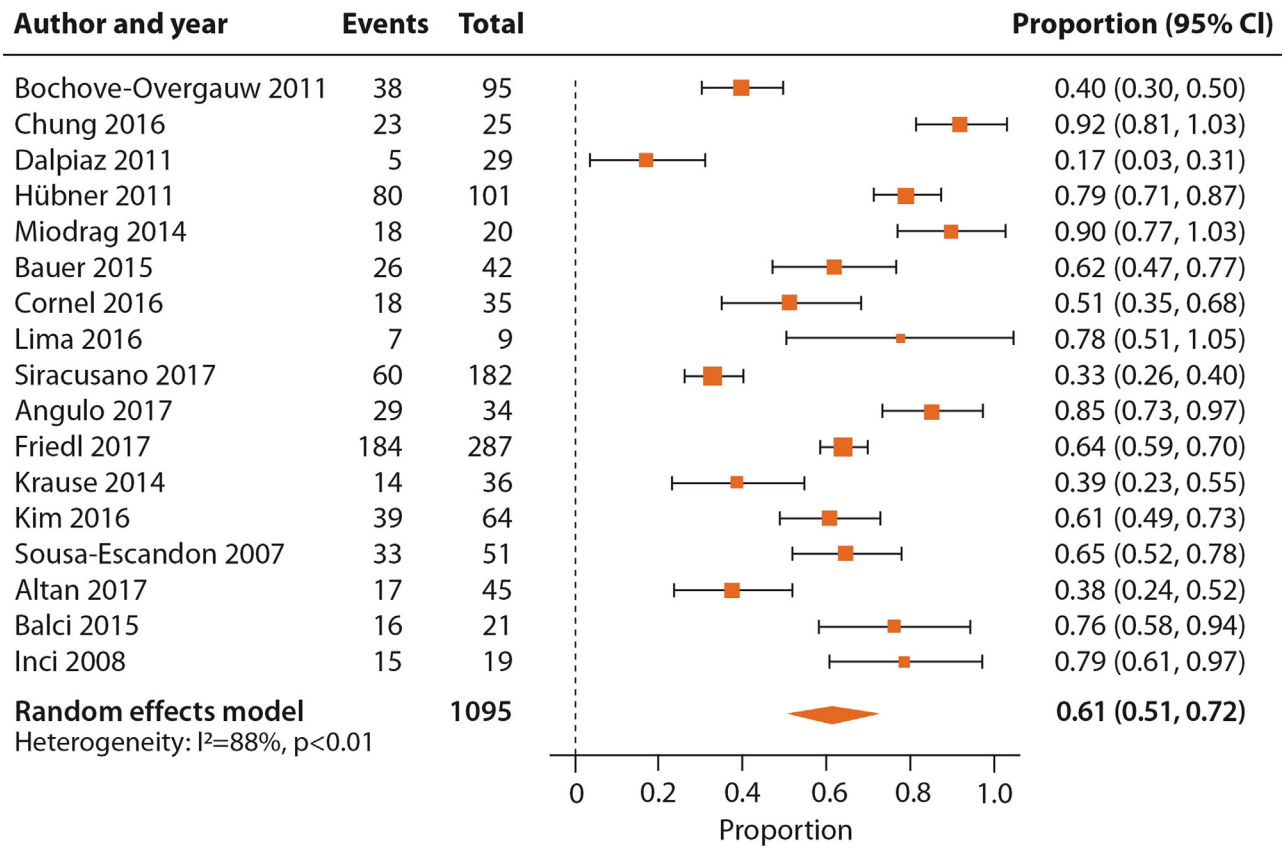


Fig. 1 – Forest plot of objective cure rates (proportions with 95% confidence interval [CI]) in 17 studies involving 1095 patients treated with adjustable slings. Adapted and recalculated from Meisterhofer et al. [10].

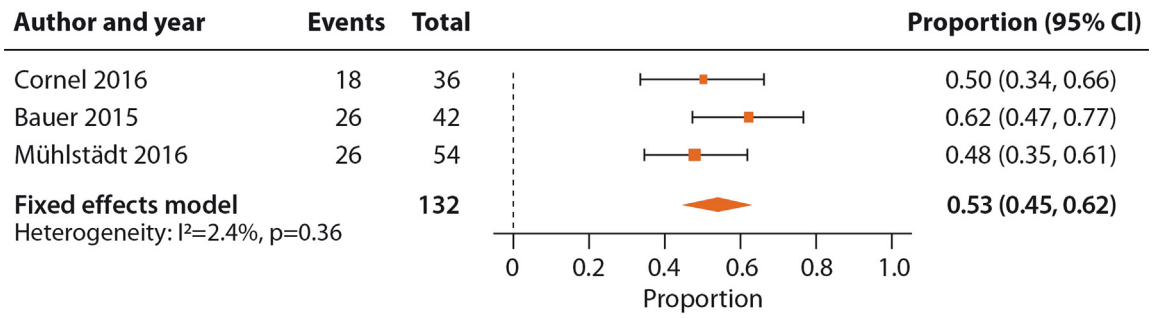


Fig. 2 – Forest plot of cure with adjustable sling (proportions with 95% confidence interval [CI]) defined as no incontinence on a 24-hour pad test in three studies with 132 patients. Adapted and recalculated from Silva et al. [11].

different types of adjustable slings. Cure was defined in eight different ways. If the results prove to be similar, this is not problematic—indeed, it suggests that the effects do not differ across incontinence severity and the type of sling, which would be a reassuring finding. Conversely, if the results differ appreciably, this degree of clinical heterogeneity presents challenges in interpreting the results.

In a forest plot of the efficacy of adjustable slings (Fig. 1), you note large differences between the point estimates with a substantial lack of overlap of the confidence intervals. You therefore conclude that chance is an unlikely explanation for the differences between the studies. Statistical tests support your inference ($p < 0.01$ and $I^2 = 88\%$).

In their methods section, the authors of this review describe possible explanations of heterogeneity: variability in the outcome definition, follow-up duration, sling type, and incontinence severity [10]. Unfortunately, their subgroup analyses revealed that none of these explanations elucidated the differences between groups. We are left with unexplained heterogeneity, which markedly undermines our confidence in the pooled cure estimate.

The second review took a different approach, restricting the analysis to only two types of adjustable slings and no incontinence on a 24-hour pad test as the definition of success with a follow-up of at least 12 months [11]. The forest plot (Fig. 2) shows similar point estimates (from 48% to 62%) with extensive overlap between confidence intervals. Statistical tests support your conclusion that the results are similar and the difference can easily be explained by chance ($p = 0.36$ and $I^2 = 2\%$).

After assessing the presence of heterogeneity in both reviews, you conclude that there is only low-quality evidence available to support the use of adjustable male slings: single-arm studies without internal comparisons and, when looking across different types of slings, large unexplained heterogeneity. Given the low-quality evidence, even though your patient is only interested in an adjustable sling, you decide to also inform him about the other interventions available. Ultimately, you and your patient face the need to decide under considerable uncertainty.

3. Conclusions

Systematic reviews and meta-analyses are often the best sources of evidence for clinical decision-making in urology. Understanding the assessment of variability in the enrolment of patients, application of interventions, and measurement of outcomes in individual studies, and in the examination of variability in results can help urologists make sense of systematic review results.

Conflicts of interest: The authors have nothing to disclose.

References

[1] Guyatt G, Rennie D, Meade MO, Cook DJ. Users' guides to the medical literature: essentials of evidence-based clinical practice. ed. 3. New York: McGraw-Hill Education; 2014.
[2] Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0. London, UK: The Cochrane Collaboration; 2011. www.handbook.cochrane.org.
[3] Wallis CJD, Saskin R, Choo R, et al. Surgery versus radiotherapy for clinically-localized prostate cancer: a systematic review and meta-analysis. Eur Urol 2016;70:21–30. https://doi.org/10.1016/j.eururo.2015.11.010.
[4] Kilpeläinen TP, Järvinen P, Tikkinen KAO. Randomized trials show a consistent benefit of radical prostatectomy on mortality outcomes. J Urol 2019;202:1106–8. https://doi.org/10.1097/JU.0000000000000423.
[5] Foroutan F, Guyatt G, Zuk V, et al. GRADE guidelines 28: use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. J Clin Epidemiol 2020;121:62–70. https://doi.org/10.1016/j.jclinepi.2019.12.023.
[6] Schandelmaier S, Briel M, Varadhan R, et al. Development of the instrument to assess the credibility of effect modification analyses (ICEMAN) in randomized controlled trials and meta-analyses. Can Med Assoc J 2020;192:E901–6. https://doi.org/10.1503/cmaj.200077.
[7] Kilpeläinen TP, Tikkinen KAO, Guyatt GH, Vernooij RWM. Evidence-based urology: subgroup analysis in randomized controlled trials. Eur Urol Focus 2021, In press.
[8] Burkhard FC, Bosch JLHR, Cruz F, et al. EAU guideline: urinary incontinence in adults. Arnhem, The Netherlands: European Association of Urology; 2020, https://uroweb.org/guideline/urinary-incontinence.
[9] Abrams P, Constable LD, Cooper D, et al. Outcomes of a noninferiority randomised controlled trial of surgery for men with urodynamic stress incontinence after prostate surgery (MASTER). Eur Urol 2021;79:812–23. https://doi.org/10.1016/j.eururo.2021.01.024.
[10] Meisterhofer K, Herzog S, Strini KA, Sebastianelli L, Bauer R, Dalpiaz O. Male slings for postprostatectomy incontinence: a systematic review and meta-analysis. Eur Urol Focus 2020;6:575–92. https://doi.org/10.1016/j.euf.2019.01.008.
[11] Silva LAD, Simonetti R, Silva EMKD. Adjustable sling for the treatment of post-prostatectomy urinary incontinence: systematic review and meta-analysis. Einstein 2019;17:eRW4508. https://doi.org/10.31744/einstein_journal/2019RW4508.