# Experiences presenting GRADE to the Guideline Development Group on the NICE Lower Urinary tract Symptoms (LUTS) Guideline

Clare Jones

Senior Research Fellow/Project Manager

The NCGC is a governance collaboration, hosted by the RCP and funded by NICE

Royal College of Physicians | The Royal College of Surgeons of England | Royal College of General Practitioners | Royal College of Nursing | NHS National Institute for Health and Clinical Excellence

## Background

- NICE Guideline on The Management of lower urinary tract symptoms (LUTS) in men. National Clinical Guideline Centre May 2010.
- LUTS in men is one of the first official clinical guidelines developed by NICE to pilot using GRADE.
- NICE previously used SIGN system of assessing the quality of evidence

## Aim

- To discuss the experiences and challenges of presenting GRADE quality assessment to the LUTS guideline development group (GDG)



## Methods

- Results of systematic reviews were presented to the GDG using GRADE
- This was used to develop recommendations for the guideline
- Trialled combination of different methods of presenting results with GRADE to determine best approach for our GDG to interpret

## Challenges

Multidisciplinary audience

- GDG made up of clinicians and patient representatives



## Results

- Initial presentation of GRADE with systematic review
  - Present quality assessment rating with evidence statements
  - No further information as how rating decided
- Revised presentations of GRADE
  - Present extensive details for the criteria considered when rating evidence quality of outcomes

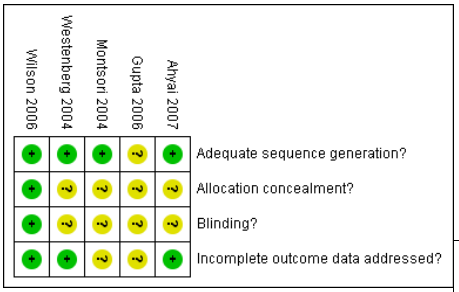
# Criteria considered for rating outcomes

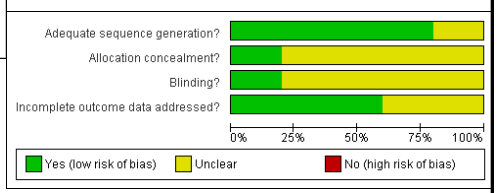
- **Limitations in design**:
  - Consider randomisation method, allocation concealment, blinding and loss to follow-up
- **Inconsistency**:
  - Consider magnitude of statistical and clinical heterogeneity between studies and explain any subgroup analysis
- **Indirectness**:
  - Does the patient population and intervention fit directly with the guideline?
- **Imprecision**:
  - How wide are the confidence intervals around the estimate of effect relative to the clinically important threshold? How many events and patients in total? Is the effect clinically significant?



# Limitations in design

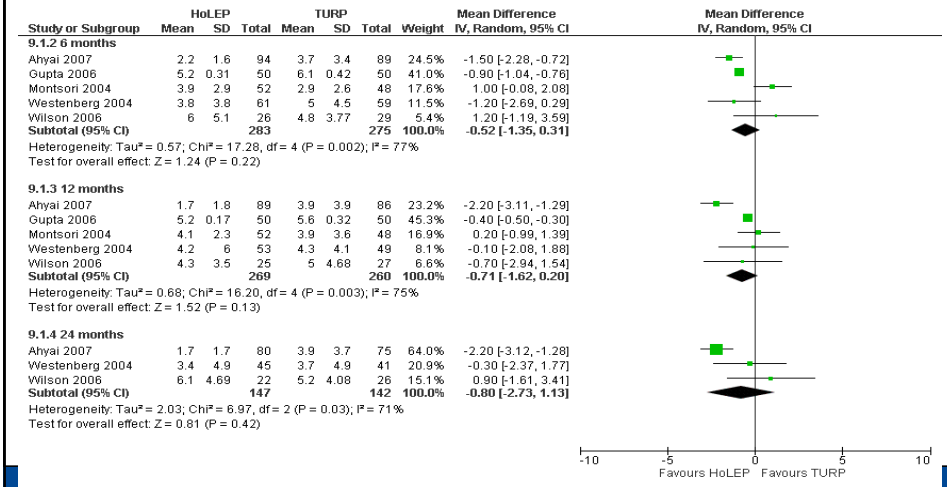
- Presented study limitations using a traffic light system to highlight the risks of bias



# Inconsistency

Presented and discussed forest plots and heterogeneity



| Study or Subgroup | HoLEP Mean | HoLEP SD | HoLEP Total | TURP Mean | TURP SD | TURP Total | Weight | Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **9.1.2 6 months** | | | | | | | | |
| Ahyai 2007 | 2.2 | 1.6 | 94 | 3.7 | 3.4 | 89 | 24.5% | -1.50 [-2.28, -0.72] |
| Gupta 2006 | 5.2 | 0.31 | 50 | 6.1 | 0.42 | 50 | 41.0% | -0.90 [-1.04, -0.76] |
| Montsori 2004 | 3.9 | 2.9 | 52 | 2.9 | 2.6 | 48 | 17.6% | 1.00 [-0.08, 2.08] |
| Westenberg 2004 | 3.8 | 3.8 | 61 | 5 | 4.5 | 59 | 11.5% | -1.20 [-2.69, 0.29] |
| Wilson 2006 | 6 | 5.1 | 26 | 4.8 | 3.77 | 29 | 5.4% | 1.20 [-1.19, 3.59] |
| **Subtotal (95% CI)** | | | **283** | | | **275** | **100.0%** | **-0.52 [-1.35, 0.31]** |
| Heterogeneity: Tau² = 0.57; Chi² = 17.28, df = 4 (P = 0.002); I² = 77% | | | | | | | | |
| Test for overall effect: Z = 1.24 (P = 0.22) | | | | | | | | |
| **9.1.3 12 months** | | | | | | | | |
| Ahyai 2007 | 1.7 | 1.8 | 89 | 3.9 | 3.9 | 86 | 23.2% | -2.20 [-3.11, -1.29] |
| Gupta 2006 | 5.2 | 0.17 | 50 | 5.6 | 0.32 | 50 | 45.3% | -0.40 [-0.50, -0.30] |
| Montsori 2004 | 4.1 | 2.3 | 52 | 3.9 | 3.6 | 48 | 16.9% | 0.20 [-0.99, 1.39] |
| Westenberg 2004 | 4.2 | 6 | 53 | 4.3 | 4.1 | 49 | 8.1% | -0.10 [-2.08, 1.88] |
| Wilson 2006 | 4.3 | 3.5 | 25 | 5 | 4.68 | 27 | 6.6% | -0.70 [-2.94, 1.54] |
| **Subtotal (95% CI)** | | | **269** | | | **260** | **100.0%** | **-0.71 [-1.62, 0.20]** |
| Heterogeneity: Tau² = 0.68; Chi² = 16.20, df = 4 (P = 0.003); I² = 75% | | | | | | | | |
| Test for overall effect: Z = 1.52 (P = 0.13) | | | | | | | | |
| **9.1.4 24 months** | | | | | | | | |
| Ahyai 2007 | 1.7 | 1.7 | 80 | 3.9 | 3.7 | 75 | 64.0% | -2.20 [-3.12, -1.28] |
| Westenberg 2004 | 3.4 | 4.9 | 45 | 3.7 | 4.9 | 41 | 20.9% | -0.30 [-2.37, 1.77] |
| Wilson 2006 | 6.1 | 4.69 | 22 | 5.2 | 4.08 | 26 | 15.1% | 0.90 [-1.61, 3.41] |
| **Subtotal (95% CI)** | | | **147** | | | **142** | **100.0%** | **-0.80 [-2.73, 1.13]** |
| Heterogeneity: Tau² = 2.03; Chi² = 6.97, df = 2 (P = 0.03); I² = 71% | | | | | | | | |
| Test for overall effect: Z = 0.81 (P = 0.42) | | | | | | | | |





# Indirectness

Does the patient population and intervention fit directly with the guideline?

# Imprecision

- Confidence intervals, minimally important differences and optimal information size values were presented to determine whether outcomes were imprecise



# Example of GRADE imprecision summary

**IPSS – optimal sample size calculation**

- Minimally important difference patient can detect is 3 points on IPSS score. Paper by Barry et al., (1995) shows for n=347 patients with AUA = slight improvement the absolute difference in IPSS score is 3 ± 5.03
- Optimal sample size calculation using these parameters gives 88 patients

| Follow up | OIS (both arms) | Total from meta-analysis | Sample size > OIS | Confidence interval cross MID | GRADE Imprecise ? |
|---|---|---|---|---|---|
| 3 mths | 88 | 213 | Yes | Yes | Yes |
| 6 mths | 88 | 165 | Yes | Yes | Yes |
| 1 yr | 88 | 283 | Yes | Yes | Yes |
| 2 yrs | 88 | 139 | Yes | Yes | Yes |
| 3 yrs | 88 | 165 | Yes | Yes | Yes |
| +5 yrs | 88 | 89 | No | Yes | Yes |

Power 80% at p=0.05 significance

Use for power calculation

http://newton.stat.ubc.ca/~rollin/stats/ssize/n2.html

# Example of presented GRADE tables

**HoLEP vs. TURP – clinical summary of findings**

| *Outcome* | *Number of Studies* | *Design* | *Limitations* | *Inconsistency* | *Directness* | *Other considerations* |
|---|---|---|---|---|---|---|
| Symptom score at 6 months | 5 | RCT | Serious limitations (a) | Serious heterogeneity (c) | No serious indirectness | No serious imprecision |
| Symptom score at 12 months | 5 | RCT | Serious limitations (a) | Serious heterogeneity (c) | No serious indirectness | No serious imprecision |
| Symptom score at 24 months | 3 | RCT | Serious limitations (a) | Serious heterogeneity (c) | No serious indirectness | No serious imprecision |

a) 4 studies did not report allocation concealment or masked outcome assessment. One study did not report randomisation method used.
b) Statistically significant heterogeneity is present.



# Example of presented GRADE tables

| *Outcome* | *Intervention* | *Control* | *Relative risk* | *Absolute effect* | *Quality* |
|---|---|---|---|---|---|
| **Mean symptom score at 6 months** | 283 | 275 | -0.52 [-1.35, 0.31] | Not applicable | LOW |
| **Mean symptom score at 12 months** | 269 | 260 | -0.71 [-1.62, 0.20] | Not applicable | LOW |
| **Mean symptom score at 24 months** | 147 | 142 | -0.80 [-2.73, 1.13] | Not applicable | LOW |

***Clinical Evidence Statements - Symptom scores***

There is no statistically significant difference between HoLEP and TURP in reducing symptom scores at 6, 12 and 24 months postoperatively. [LOW]

## Quality ratings

- Knowledge of literature on topic
- Concern regarding evidence with low quality or very low quality



## Quality ratings

- GRADE definition of quality provided to group to illustrate the difference between the GRADE and NICE previous system

## Overall quality for outcome evidence

- **High**:
  - further research is very unlikely to change our confidence in the estimate of effect
- **Moderate**:
  - further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- **Low**:
  - further research is ***very*** likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- **Very low**:
  - any estimate of effect is very uncertain



## Key Points

- Critical to include the right level of detail
- Transparency of process important
- Ensure the GDG understand the process and support the quality rating which underpins the decisions and recommendations made

## LUTS guideline

- Published May 2010
- http://guidance.nice.org.uk/CG97

THANK YOU!