Reproducibility of Renal Split Function

preliminary results on a worldwide study

Anton Staudenherz
G. Dobrozemsky, Ch. Krafl, M. Samat, P.M. Schaflairich, H. Sinzinger

clinical decision making

- paediatric nephro-urology
  - interdisciplinary meetings
- the more we learn and know
  - knowledge in diagnostics and therapeutics
- the less invasive the therapy decisions are
  - wait and see instead of surgery

the index patient

- 4 weeks old: 4.1 kg, 53cm; hydronephrosis
  - baseline study
- 9 weeks old: 5.9 kg, 60cm; hydronephrosis
  - follow-up study – significant functional impairment?
follow-up – right decision?

- 4 weeks old: 4.1 kg; 53cm; hydronephrosis
  - baseline study
- 9 weeks old: 5.9 kg; 60cm; hydronephrosis
  - follow up study – significant decrease in function?
- what to do???
  - decision: wait and see – was it the right one???

background

- radioisotopic renography – first introduced in 1956
- widely accepted quantitative parameter – split function (basis for clinical decision making)
- why a round-robin test?
  - quality control
  - economic decisions
  - quality assurance
  - are my results correct

design

- round – robin test for kidney analysis software
- 3 patient (commonly) and 2 phantoms
- dynamic renal studies were anonymized and distributed
- participants were asked to assess split function (SF) using their own routine software
- in a simple form the measurements of SF should be reported and sent back
• 3 patients and 2 simulations – rather small series – statistical power?
• insufficient acquisition parameter (zoom)
• simulation studies are too easy – should not be part of a study
• no gold standard
• matrix size
• problems due to conversion
how many countries?

- 218 countries worldwide
- 192 countries approved from the UN

feedback

- assuming 82 countries with Nuclear Medicine Departments (= list of WFNMB Members)
- we got feedback from 21 countries from all continents - reflecting a return rate 26%
statistical methods

• characteristics – descriptive analysis
  • mean, standard deviation

• analysis using Iso-Guides 43-1 for round robin tests
  • outlier identified using Grubbs-Test z-score calculation

\begin{align*}
  z_i = \frac{x_i - \bar{X}}{\sigma} \\
  \bar{X} \ldots test \ value \\
  \sigma \ldots \text{standardized uncertainty} \\
  u_{\text{vor}} \ldots \text{measurement uncertainty} \\
  \sigma = 0.5 \times u_{\text{vor}} \\
  u_{\text{vor}} = \sqrt{u_{\text{ref}}^2 + u_{\text{lab}}^2}
\end{align*}

results and aims

• from 126 calculated split functions, only 62% are acceptable (tolerance level within ±10%)
• our sample is representative
• each center will get a feedback
• high impact publication
routine tracer
- 76% MAG3
- 21% DTPA
- 3% Iodine-Hippurate

distribution

tolerance limit < 10%
relative error below 10%
- study_01: 89%
- study_02: 22%
- study_03: 62%
- study_04: 100%
- study_05: 36%

\[ E_{\text{rel}} = \frac{|m - w_{\text{ref}}|}{w_{\text{ref}}} \]
z-score

<table>
<thead>
<tr>
<th>z-score</th>
<th>≤ 1</th>
<th>≥ 2</th>
<th>≥ 3</th>
<th>≥ 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>result</td>
<td>good</td>
<td>acceptable</td>
<td>questionable</td>
<td>insufficient</td>
</tr>
</tbody>
</table>

- study_01: 56% ≤ 1, 27% ≥ 2, 3% ≥ 3, 14% ≥ 4
- study_02: 16% ≤ 1, 11% ≥ 2, 8% ≥ 3, 65% ≥ 4
- study_03: 41% ≤ 1, 24% ≥ 2, 5% ≥ 3, 30% ≥ 4
- study_04: 61% ≤ 1, 25% ≥ 2, 11% ≥ 3, 3% ≥ 4
- study_05: 28% ≤ 1, 22% ≥ 2, 33% ≥ 3, 17% ≥ 4

z-score overall

- ≥ 1: good result
- ≥ 2: acceptable result
- ≥ 3: questionable result
- ≥ 4: insufficient result

Problem

- the diseased kidney
  - split function < 35%
• what might be the confounding factor for unacceptable results?
• so far the variance of the preliminary results is surprising - is SF really a robust parameter?
• think positive – it is worth to be further evaluated (international standardization?)