Clinical utility of prediction models for ovarian tumor diagnosis: a decision curve analysis

L. Wynants, J. Verbakel, S. Van Huffel, D. Timmerman, B. Van Calster

MEMTAB, 19 July 2016
Diagnostic models for ovarian cancer

• Women with suspicious masses in the adnexa
• Pre-operative diagnosis required to determine the clinical pathway (general vs. high-volume hospital, general surgeon or gynaecologist vs. gynaecological oncologist)
• Ultrasound scan preferred method
• Experience required
• Diagnostic models
## Diagnostic models for ovarian cancer

### Develop model

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMI (Jacobs et. al., 1990)</td>
<td>Menopausal status, CA125, ultrasound variables</td>
<td>A numerical score, cut-off 200</td>
</tr>
<tr>
<td>ROMA (Moore et al., 2009)</td>
<td>CA125, HE4 and menopausal status</td>
<td>Risk</td>
</tr>
<tr>
<td>IOTA LR2 (Van Holsbeke et al., 2009)</td>
<td>Age, ultrasound variables</td>
<td>Risk</td>
</tr>
<tr>
<td>IOTA Simple Rules Risks (Timmerman et al., 2016)</td>
<td>Ultrasound variables, centre type</td>
<td>Risk</td>
</tr>
<tr>
<td>IOTA ADNEX (Van Calster et al., 2015)</td>
<td>Age, CA125, ultrasound variables, centre type</td>
<td>Risk for various tumor types</td>
</tr>
</tbody>
</table>

### Test model in new data

- **Develop model**
- **Test model in new data**
Diagnostic models for ovarian cancer

Develop model

Test model in new data

<table>
<thead>
<tr>
<th>Model</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMI</td>
<td>Jacobs et. al., 1990</td>
</tr>
<tr>
<td>ROMA</td>
<td>Moore et al., 2009</td>
</tr>
<tr>
<td>IOTA LR2</td>
<td>Van Holsbeke et al., 2009</td>
</tr>
<tr>
<td>IOTA Simple Rules Risks</td>
<td>Timmerman et al., 2016</td>
</tr>
<tr>
<td>IOTA ADNEX</td>
<td>Van Calster et al., 2014</td>
</tr>
</tbody>
</table>

- Women with adnexal mass scheduled for surgery
- Histology
Diagnostic models for ovarian cancer

Develop model

Test model in new data

<table>
<thead>
<tr>
<th>Model</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMI</td>
<td>(Jacobs et. al., 1990)</td>
</tr>
<tr>
<td>ROMA</td>
<td>(Moore et al., 2009)</td>
</tr>
<tr>
<td>IOTA LR2</td>
<td>(Van Holsbeke et al., 2009)</td>
</tr>
<tr>
<td>IOTA Simple Rules Risks</td>
<td>(Timmerman et al., 2016)</td>
</tr>
<tr>
<td>IOTA ADNEX</td>
<td>(Van Calster et al., 2014)</td>
</tr>
</tbody>
</table>

- Women with adnexal mass scheduled for surgery
- Histology

- Dataset 1: 2009-2012, 18 centres from six European countries, oncology centres and non-oncology centres, n=2403 (980 malignant)
Diagnostic models for ovarian cancer

Develop model | Test model in new data

<table>
<thead>
<tr>
<th>Model</th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMI (Jacobs et. al, 1990)</td>
<td>2009-2012, 18 centres from six European countries, oncology centres and non-oncology centres, n=2403 (980 malignant)</td>
<td>Women with adnexal mass scheduled for surgery</td>
</tr>
<tr>
<td>ROMA (Moore et al., 2009)</td>
<td></td>
<td>Histology</td>
</tr>
<tr>
<td>IOTA LR2 (Van Holsbeke et al., 2009)</td>
<td></td>
<td>Dataset 1: 2009-2012, 18 centres from six European countries, oncology centres and non-oncology centres, n=2403 (980 malignant)</td>
</tr>
<tr>
<td>IOTA Simple Rules Risks (Timmerman et al., 2016)</td>
<td></td>
<td>Dataset 2: 2005-2009, Belgian oncology centre, n=360 (144 malignant)</td>
</tr>
<tr>
<td>IOTA ADNEX (Van Calster et al., 2014)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Women with adnexal mass scheduled for surgery
Histology
Dataset 1: 2009-2012, 18 centres from six European countries, oncology centres and non-oncology centres, n=2403 (980 malignant)
Dataset 2: 2005-2009, Belgian oncology centre, n=360 (144 malignant)
## Diagnostic models for ovarian cancer

**Develop model**

<table>
<thead>
<tr>
<th>Model</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMI</td>
<td>(Jacobs et. al., 1990)</td>
</tr>
<tr>
<td>ROMA</td>
<td>(Moore et al., 2009)</td>
</tr>
<tr>
<td>IOTA LR2</td>
<td>(Van Holsbeke et al., 2009)</td>
</tr>
<tr>
<td>IOTA Simple Rules Risks</td>
<td>(Timmerman et al., 2016)</td>
</tr>
<tr>
<td>IOTA ADNEX</td>
<td>(Van Calster et al., 2014)</td>
</tr>
</tbody>
</table>

**Test model in new data**

- Women with adnexal mass scheduled for surgery
- Histology
- Discrimination
- Calibration
- Clinical utility
  - Is the model useful to guide decision-making?
Net Benefit

- The ‘harm-to-benefit ratio’ of treating a false positive (FP) versus a true positive (TP). (denoted by “w”)

- Net Benefit = \[ \frac{\text{number of true positives} - w \times \text{number of false positives}}{\text{total number of observations}} \]

- \( w = \frac{t}{1-t} \), where \( t \) is the risk threshold used to classify patients as positive (events) or negative (non-events).

- Risk thresholds between 5% and 50%.
  - 5% (odds 5:95): equivalent to accepting up to 19 false positives (FP) per true positive (TP)
  - 50% (odds 50:50): equivalent to accepting up to 1 false positives (FP) per true positive (TP).
Net Benefit

- Net Benefit = \( \frac{\text{number of true positives} - w \times \text{number of false positives}}{\text{total number of observations}} \)

- Compare models to each other and to the default strategies: treat (refer) all and treat (refer) none.

- The higher the NB, the more clinical utility.
Net Benefit

- Net Benefit = \( \frac{\text{number of true positives} - w \times \text{number of false positives}}{\text{total number of observations}} \)

- Compare models to each other and to the default strategies: treat (refer) all and treat (refer) none.

- The higher the NB, the more clinical utility.
  - We expressed the difference between models as gain in 'net sensitivity' (i.e., sensitivity for a constant specificity, \( \Delta \text{NB/prevalence} \)).
Net Benefit

- Net Benefit = \( \frac{\text{number of true positives} - w \times \text{number of false positives}}{\text{total number of observations}} \)

- Compare models to each other and to the default strategies: treat (refer) all and treat (refer) none.

- The higher the NB, the more clinical utility.
  - gain in ‘net sensitivity’ (i.e., sensitivity for a constant specificity, \( \Delta NB / \text{prevalence} \)).
  - net amount of avoided false positives per hundred patients (i.e., avoided false positives for a constant number of true positives, \( 100 \times (1-t)/(t) \times \Delta NB \)).
  - 95% bootstrap confidence intervals
Results dataset 1
Results dataset 1

Srrisks vs LR2:
Sens: 4% [3% to 6%]
Avoided FP: 16 [10 to 23]
Results dataset 1

Srrisks vs ADNEX:
Sens: 0% [-1% to 2%]
Avoided FP: 1 [-3 to 6]
Results dataset 1

Oncology centers

Other centers
Results dataset 1

**Premenopausal**

- None
- All
- RMI
- LR2
- SRRisks
- ADNEX

**Postmenopausal**

- Net benefit
- Risk threshold (%)
- Number of FP for 1 TP
Results dataset 2
Results dataset 2
Conclusion

- NB supersedes discrimination and calibration to quantify the clinical utility of prediction models.
- ADNEX and SRRisks has shown potential to improve triage of patients with ovarian masses.
- RMI and ROMA have no utility for clinical decision making at low risk thresholds, and lower utility than ADNEX and SRRisks at higher risk thresholds.
Thank you

Questions? Comments?