What is a reasonable comparator test to run against for new tests

Carlo Senore

Epidemiology and screening unit – CPO

University hospital Città della salute e delal Scienza, Turin, Italy

Chairs:
Bob Bresalier, Carlo Senore, Graeme Young
Conflicts

None to declare
**Comparator test**

**Guiding principles 2016**

*gFOBT* is the **minimum standard** (29-47% sensitivity for CRC; 87%-98% specificity)

*FIT* also **acceptable** (superior to *gFOBT*)

---

**FIT** New standard:

- Improved sensitivity for CRC and better capacity to detect (advanced) adenomas
- Lower interval cancer rate
- Repeated testing improves detection
- Preliminary reports from population based studies are suggestive for the effectiveness of FIT in reducing CRC incidence and mortality

Comparator test

USPSTF suggests (single test) (Bibbins-Domingo et al., 2016)

• acceptable sensitivity for CRC (all stages) to be at least 70%

• acceptable specificity to be at least 90%.

Desirable standards have not been set for advanced adenomas

FIT (cutoff 20 µg Hb/gr.):

• sensitivity for CRC (all stages): 75% (61-86%)

• specificity: 95% (92-98%) (Imperiale et al., 2019)
FIT as a comparator

Different FIT systems available

**Qualitative and quantitative** methods

Variable **number of samples**, even if most programs require a single sample

Variability in the choice of the **positivity cut-off**

FIT accuracy is influenced by test characteristics, cut-off, protocol

**Need to set a standard**

**Quantitative FIT methods**

**Single sample**
FIT as a comparator

Different quantitative methods are showing different performance characteristics

Comparing FIT methods performance at identical positivity rates

(Grobbee et al. 2016; Passamonti et al. 2018)
FIT as a comparator

9 - A screening test with an adjustable threshold (or algorithm) for test positivity enables a program to choose the test accuracy parameters that best achieve desired program outcomes.

FIT : positivity threshold ideally low, in studies with colonoscopy follow-up only for subjects testing positive at either test, to allow simulating comparisons for different cut-off levels and decreasing positivity rates

New test: Cut-off needs to be flexible -> performance should be measured simulating different cut-off levels, focusing on the positivity range of established population based programs
Cross-sectional evaluation

**Paired testing within ongoing screening programs**

- immediate assessment of
- all subjects who are positive on the comparator test
- all those testing positive with the new-test only
Prospective evaluation

Comments from the consensus rounds

Requiring a new test to be tested for multiple rounds is ideal, but it seems a very high standard

Tests need to be brought to the market and then the subsequent evaluation needs to be done
Prospective evaluation

Detection of early stage lesions is a goal and therefore stage shift may be considered as a plausible surrogate measure, but screening may increase the detection of indolent lesions.

Change in the absolute risk of late stage diagnosis would seem a more valid measure.

The new test, apart from detecting new lesions compared to the reference, might miss lesions detected by the reference test.

- to assess the clinical significance of these discordant cases prospective comparative studies would be necessary.
Prospective evaluation

Impact of screening programme using the faecal immunochemical test on stage of colorectal cancer: Results from the IMPATTO study

Massimo Vicentini, Manuel Zorzi, Emanuela Bovo, Pamela Mancuso, Marco Zappa, Gianfranco Manneschi, Lucia Mangone, and Paolo Giorgi Rossi, Colorectal Cancer Screening IMPATTO study working group

Relative Incidence by stage

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II-III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not invited</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Attendees 1 episode</td>
<td>4.6 (4.2 - 5.1)</td>
<td>1.4 (1.3 - 1.5)</td>
<td>0.7 (0.6 - 0.9)</td>
</tr>
<tr>
<td>Attendees following episode</td>
<td>1.4 (1.2 - 1.6)</td>
<td>0.8 (0.7 - 0.9)</td>
<td>0.3 (0.1 - 0.4)</td>
</tr>
</tbody>
</table>
Interval CRCs

Prospective evaluation

Table 1. Comparison of Four Fecal Occult-Blood Tests in 8104 People.

<table>
<thead>
<tr>
<th>TEST AND FINDING</th>
<th>NEOPLASMS DETECTED</th>
<th>TRUE POSITIVE TEST</th>
<th>FALSE POSITIVE TEST</th>
<th>TRUE NEGATIVE TEST</th>
<th>FALSE NEGATIVE TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no.</td>
<td>no. of patients (%)</td>
<td>no.</td>
<td>no. of patients (%)</td>
<td>no. of patients (%)</td>
</tr>
<tr>
<td>Hemoccult II (8065 screened, 198 [2.5%] positive)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carcinoma</td>
<td>35</td>
<td>13 (0.2)</td>
<td>185 (2.3)</td>
<td>7845 (97.3)</td>
<td>22 (0.3)</td>
</tr>
<tr>
<td>Polyp $\geq$ 1 cm</td>
<td>107</td>
<td>33 (0.4)</td>
<td>152 (1.9)</td>
<td>7771 (96.8)</td>
<td>74 (0.9)</td>
</tr>
<tr>
<td>Combined</td>
<td>142</td>
<td>46 (0.6)</td>
<td>152 (1.9)</td>
<td>7771 (96.4)</td>
<td>96 (1.2)</td>
</tr>
<tr>
<td>Hemoccult II Sensa (7904 screened, 1073 [13.6%] positive)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carcinoma</td>
<td>34</td>
<td>27 (0.3)</td>
<td>1046 (13.2)</td>
<td>6824 (86.3)</td>
<td>7 (0.1)</td>
</tr>
<tr>
<td>Polyp $\geq$ 1 cm</td>
<td>105</td>
<td>72 (0.9)</td>
<td>974 (12.4)</td>
<td>6791 (86.3)</td>
<td>33 (0.4)</td>
</tr>
<tr>
<td>Combined</td>
<td>139</td>
<td>99 (1.3)</td>
<td>974 (12.3)</td>
<td>6791 (85.9)</td>
<td>40 (0.5)</td>
</tr>
<tr>
<td>HemeSelect (7493 screened, 440 [5.9%] positive)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carcinoma</td>
<td>32</td>
<td>22 (0.3)</td>
<td>418 (5.6)</td>
<td>7043 (94.0)</td>
<td>10 (0.1)</td>
</tr>
<tr>
<td>Polyp $\geq$ 1 cm</td>
<td>102</td>
<td>68 (0.9)</td>
<td>350 (4.7)</td>
<td>7009 (93.9)</td>
<td>34 (0.5)</td>
</tr>
<tr>
<td>Combined</td>
<td>134</td>
<td>90 (1.2)</td>
<td>350 (4.7)</td>
<td>7009 (93.5)</td>
<td>44 (0.6)</td>
</tr>
<tr>
<td>Combination (Hemoccult II Sensa and HemeSelect)</td>
<td>(7847 screened, 233 [3.0%] positive)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carcinoma</td>
<td>32</td>
<td>21 (0.3)</td>
<td>212 (2.7)</td>
<td>7603 (96.9)</td>
<td>11 (0.1)</td>
</tr>
<tr>
<td>Polyp $\geq$ 1 cm</td>
<td>102</td>
<td>51 (0.7)</td>
<td>161 (2.1)</td>
<td>7552 (96.6)</td>
<td>51 (0.7)</td>
</tr>
<tr>
<td>Combined</td>
<td>134</td>
<td>72 (0.9)</td>
<td>161 (2.1)</td>
<td>7552 (96.2)</td>
<td>62 (0.8)</td>
</tr>
</tbody>
</table>

*For carcinoma the percentages are based on the number of patients screened. For polyps the percentages are based on the number of patients screened minus the number with carcinoma detected on colonoscopy.

Allison J et al. NEJM 1996

As long as screening needs to be repeated we need to get prospective information supporting the decision about the **optimal screening interval**
Prospective evaluation

A separate step in the approval process

The test could be approved for individual use after rigorous head-to-head comparison and for use in organized programs after rigorous cost-effectiveness evaluation over multiple rounds.
New imaging methods: TC colonography, colon capsule endoscopy

These tests would be offered at long intervals

Would sigmoidoscopy still represent a valid comparator in these cases, as indicated in the 2016 guiding principles?

Sigmoidoscopy: comparator for the detection of pre-invasive lesions
The context for comparative evaluation

Learning programs: ongoing programs as a platform for research/introducing new technologies

Biobanking within existing programs

Thank you for your attention

carlo.senore@cpo.it