High proportion of high quality randomized clinical trials conducted by the NCI are negative or inconclusive

Ambuj Kumar, Heloisa Soares

and

Benjamin Djulbegovic

Moffitt Cancer Center & Research Institute

at the

University of South Florida
Phase III Randomized clinical trials (RCT) remain the primary mean for development of new treatments for the prevention and cure of cancer.

However, sometimes a RCT fails to show a significant difference between the experimental and the control treatments.
Objective

True negative vs. false-negative results?

• Is the new intervention truly not effective, i.e.
  evidence of absence of treatment effect
  or
• The trial's results were inconclusive, i.e.
  absence of evidence of treatment effect

Altman, DG et.al. in *BMJ* 1995;311:485 (19 August)
Defining true negative or inconclusive

• True negative
  - if the effect size and the 95% CIs were entirely outside the pre-determined limit of equivalence

• Inconclusive
  - if the 95% CIs crossed the line of no effect and one or both limits of pre-determined equivalence
Interpretation

Insufficient evidence to confirm or exclude if experimental treatment is better than the standard or vice-versa.

Outcomes statistically significant favoring innovation.

Statistically significant difference, unclear if it's important to patients.

Statistically significant difference, not important to patients.

Important difference.

Characteristics of Confidence interval

Pre-defined limits of equivalence

Inconclusive

True Negative (excluding benefit from experimental treatment)

Outcomes statistically significant favoring standard

Line of no effect (relative risk, odds ratio=1, risk difference=0)

Adapted from Alderson, P. BMJ 2004;328:476-477
Methods

All consecutive phase III RCTs conducted by three NCI sponsored Cooperative Groups were reviewed (protocols and final publications)

<table>
<thead>
<tr>
<th>Cooperative group</th>
<th>No. of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation Therapy Oncology Group (RTOG)</td>
<td>38</td>
</tr>
<tr>
<td>Children's Oncology Group (ChOG)</td>
<td>91</td>
</tr>
<tr>
<td>Gynecologic Oncology Group (GOG)</td>
<td>25</td>
</tr>
</tbody>
</table>

All consecutive trials from 1955-2000
Why NCI-sponsored cooperative group RCTs?

- NCI-sponsored COGs conduct all the publicly funded RCTs in the USA
- All COG research protocols pass a rigorous peer-review process.
3 NCI sponsored cooperative group trials included in the review
(Radiation Therapy Oncology Group, Children’s Oncology Group and Gynecologic Oncology Group)

N=261 (~50,000 patients)

Outcome statistically significant
36% (93/261)

- Favoring innovation: 70% (65/93)
- Favoring Standard: 30% (28/93)

Outcome statistically not significant
64% (168/261)*
*(data available for 148/168 studies)

- True negative: 66% (98/148)
- Inconclusive: 34% (50/148)
### Meta Analysis – inconclusive trials

**Primary end point: Survival**

<table>
<thead>
<tr>
<th>Study</th>
<th>Deaths/Patients</th>
<th>Statistics (O-E)</th>
<th>O.R. &amp; 95% CI (Innovation : Standard)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Innovation</td>
<td>Standard</td>
<td>Var.</td>
</tr>
<tr>
<td>COG 101/143</td>
<td>11/160</td>
<td>17/156</td>
<td>-2.9</td>
</tr>
<tr>
<td>COG 191p</td>
<td>32/122</td>
<td>12/59</td>
<td>1.2</td>
</tr>
<tr>
<td>COG 681/7898a</td>
<td>9/84</td>
<td>5/37</td>
<td>0.7</td>
</tr>
<tr>
<td>COG 681/7898c</td>
<td>64/88</td>
<td>62/83</td>
<td>-0.7</td>
</tr>
<tr>
<td>COG 7208b</td>
<td>27/90</td>
<td>23/87</td>
<td>2.1</td>
</tr>
<tr>
<td>COG 7409</td>
<td>11/19</td>
<td>7/25</td>
<td>3.8</td>
</tr>
<tr>
<td>COG 7422</td>
<td>4/40</td>
<td>4/43</td>
<td>0.2</td>
</tr>
<tr>
<td>COG 7721</td>
<td>41/55</td>
<td>38/59</td>
<td>6.4</td>
</tr>
<tr>
<td>COG 8725</td>
<td>10/80</td>
<td>5/81</td>
<td>0.1</td>
</tr>
<tr>
<td>COG 8821/22</td>
<td>55/115</td>
<td>75/117</td>
<td>-9.4</td>
</tr>
<tr>
<td>COG 9239</td>
<td>59/85</td>
<td>62/66</td>
<td>-2.5</td>
</tr>
<tr>
<td>COG 943</td>
<td>18/28</td>
<td>25/30</td>
<td>10.8</td>
</tr>
<tr>
<td>COG 944</td>
<td>27/35</td>
<td>29/35</td>
<td>2.2</td>
</tr>
<tr>
<td>RTOG 7102</td>
<td>63/73</td>
<td>57/68</td>
<td>-0.2</td>
</tr>
<tr>
<td>RTOG 7103</td>
<td>31/55</td>
<td>30/54</td>
<td>-0.3</td>
</tr>
<tr>
<td>RTOG 7104</td>
<td>61/71</td>
<td>59/68</td>
<td>-0.8</td>
</tr>
<tr>
<td>RTOG 7105</td>
<td>79/140</td>
<td>63/132</td>
<td>8.5</td>
</tr>
<tr>
<td>RTOG 7301</td>
<td>265/277</td>
<td>97/102</td>
<td>2.3</td>
</tr>
<tr>
<td>RTOG 7610</td>
<td>140/163</td>
<td>104/134</td>
<td>6.9</td>
</tr>
<tr>
<td>RTOG 7907</td>
<td>44/63</td>
<td>28/39</td>
<td>-0.9</td>
</tr>
<tr>
<td>RTOG 7921</td>
<td>26/26</td>
<td>23/23</td>
<td>0.0</td>
</tr>
<tr>
<td>RTOG 8403</td>
<td>80/93</td>
<td>79/94</td>
<td>5.8</td>
</tr>
<tr>
<td>RTOG 8522</td>
<td>66/83</td>
<td>65/86</td>
<td>-0.4</td>
</tr>
<tr>
<td>RTOG 8524</td>
<td>85/99</td>
<td>85/94</td>
<td>-0.4</td>
</tr>
<tr>
<td>RTOG 9104</td>
<td>175/211</td>
<td>165/210</td>
<td>5.0</td>
</tr>
<tr>
<td>GOG 20</td>
<td>30/75</td>
<td>39/81</td>
<td>-0.4</td>
</tr>
<tr>
<td>GOG 23</td>
<td>96/113</td>
<td>51/61</td>
<td>1.4</td>
</tr>
<tr>
<td>GOG 24</td>
<td>74/135</td>
<td>81/148</td>
<td>5.0</td>
</tr>
<tr>
<td>GOG 56</td>
<td>67/137</td>
<td>89/157</td>
<td>11.5</td>
</tr>
<tr>
<td>GOG 95</td>
<td>37/119</td>
<td>40/110</td>
<td>-3.5</td>
</tr>
<tr>
<td>GOG 97</td>
<td>176/223</td>
<td>164/235</td>
<td>5.8</td>
</tr>
</tbody>
</table>
Why there were so many inconclusive studies?
Critical components of a RCT

\[ \alpha \text{ (usually 0.05) } \quad \beta \text{ (usually 0.2) } \]

\[ \alpha, \beta \text{ usually fixed} \]

Sample size = \( N_{\text{inn}} + N_{\text{std}} = N_{t} \text{ (total)} \)

\[ \Delta \text{ Effect size (expected difference)} \]
Results

• Quality of trials was high.

• 70% (103/148) of the studies had undertaken a pre-trial power analysis.

• The investigators chose to detect difference in primary outcomes between competing treatments ranging from 9% to > 100%.
Distribution of expected difference in primary outcome (as stated in research protocols)

- 9-10%: 6 studies
- 11-15%: 18 studies
- 16-20%: 29 studies
- 21-25%: 12 studies
- 26-38%: 8 studies
- 39-50%: 16 studies
- 51-75%: 3 studies
- >75-100%: 11 studies

Expected difference in primary outcome (a priori, 103 studies)
Planned accrual versus actually accrued (inconclusive studies only)
Expectation bias – the culprit?

Expected versus observed difference in primary outcome (inconclusive studies)

-20.00%
-10.00%
0.00%
10.00%
20.00%
30.00%
40.00%

Observed
Expected

Finding in the same direction (not as expected)
Finding in the reverse direction (opposite to expected)
Expectation bias – the culprit?

Finding in the same direction (not as expected)

Finding in the reverse direction (opposite to expected)

Expected versus observed difference in primary outcome (negative studies)
Conclusion

• Even high-quality RCTs conducted by prestigious institutions and respected research groups often produce inconclusive or negative findings

• That is, results that are statistically consistent with both, absence and presence of a benefit
Unrealistic expectations in treatment effect

- Investigators rarely, if ever, provided a rationale for determination of the chosen effect size.
Conclusions

- Unrealistic expectations in treatment effect may hamper advancements in medicine.
- Making investigators aware of their unrealistic expectations may result in designing more realistic studies.
  - Which can optimize the chances of discovery of small but worthwhile treatment effects.
- Precious resources were wasted.
- Patients participated in unnecessary trials.
  - Breach of contract with patients.
Thank you

This research was supported by the NIH/ORI

(PI: Dr. Ben Djulbegovic)

Grant #: 1 R01 NS044417-01 and 1 R01 NS052956-01