Lessons learnt from scoring adjuvant colon cancer trials and meta-analyses using the ESMO-Magnitude of Clinical Benefit Scale V.1.1

Daan Geert Knapen,1 Nathan I Cherny,2 Panagiota Zygoura,3 Nicola Jane Latino,4 Jean-Yves Douillard,4 Urania Dafni,5,6 Elisabeth G E de Vries,1 Derk Jan de Groot 1

ABSTRACT

Background Form 1 of the European Society for Medical Oncology-Magnitude of Clinical Benefit Scale (ESMO-MCBS) serves to grade therapies with curative intent. Hitherto only few trials with curative intent have been field tested using form 1. We aimed to evaluate the applicability of the scale and to assess the reasonableness of the generated scores in early colon cancer, in order to identify shortcomings that may be rectified in future amendments.

Methods Adjuvant studies were identified in PubMed, Food and Drug Administration and European Medicines Agency registration sites, as well as ESMO and National Comprehensive Cancer Network guidelines. Studies meeting inclusion criteria were graded using form 1 of the ESMO-MCBS V.1.1 and field tested by ESMO Colorectal Cancer Faculty. Shortcomings of the scale were identified and evaluated.

Results Eighteen of 57 trials and 7 out of 14 meta-analyses identified met criteria for ESMO-MCBS V.1.1 grading. In stage III colon cancer, randomised clinical trials and meta-analyses of modulated 5-fluorouracil (5-FU) based chemotherapy versus surgery scored ESMO-MCBS grade A and randomised controlled trials (RCTs) and meta-analyses comprising oxaliplatin added to this 5-FU backbone showed a more modest additional overall survival benefit (grade A and B). For stage II colon cancer, the findings are less consistent. The fluoropyrimidine trials in stage II were graded ‘no evaluable benefit’ but the most recent meta-analysis demonstrated a 5.4% survival advantage after 8 years follow-up (grade A). RCTs and a meta-analysis adding oxaliplatin demonstrated no added benefit. Exploratory toxicity evaluation and annotation was problematic given inconsistent toxicity reporting and limited results of late toxicity. Field testers (n=37) reviewed the scores, 25 confirmed their reasonableness, 12 found them mostly reasonable. Moreover, they identified the inability of crediting improved convenience in non-inferiority trials as a shortcoming.

Conclusion Form 1 of the ESMO-MCBS V.1.1 provided very reasonable grading for adjuvant colon cancer studies.

INTRODUCTION

Colorectal cancer is the third most common tumour in men, the second in women and second place in cancer-related cause of death in the world.1 Mortality has declined over the years for several reasons, including colorectal cancer screening and more effective systemic therapies in both the adjuvant setting and metastatic disease.

Adjuvant therapies for colon cancer have evolved over the past 40 years. Early studies failed to show overall survival (OS) benefit of single agent therapy including 5-fluorouracil (5-FU) monotherapy compared with surgery.
adjuvant therapy consists of a 5-fluorouracil (5-FU) or UFT, uracil and tegafur. Metastatic treatment of advanced and incurable cancer using forms 2a, b, c and 3 of the ESMO-MCBS has been extensively field tested and reviewed for reasonableness, hitherto only 13 trials with curative intent, including adjuvant therapies, had been field tested using form 1.16 The main purpose was to evaluate the applicability of the scale in adjuvant colon cancer trials and further assess the reasonableness of the generated scores, in order to identify shortcomings that may be rectified in future amendments.

**METHODS**

Randomised controlled trials (RCTs) and meta-analyses in the adjuvant treatment of stage II/III colon cancer, published since the review of negative studies by Buyse et al8 up to September 2019 were identified. Data were collected by electronic searches of PubMed (medical headings “colonic neoplasms” OR “colorectal neoplasms”, and the text words “adjuvant therapy” OR “adjuvant chemotherapy” OR “early colon cancer”) and by a manual review of Food and Drug Administration and European Medicines Agency registration sites and ESMO and National Comprehensive Cancer Network guidelines.3 17 Reference lists of included studies were also analysed.

We included also trials that investigated regimens, which are currently seen as obsolete, to ensure the most comprehensive overview of the treatment of early colon cancer over three decades. These obsolete regimens often serve as the control arm in newer trials. Furthermore, scoring these older trials might give valuable information regarding the applicability of the scale and identify possible shortcomings. Trials investigating adjuvant treatment regimens that resulted in only negative results were excluded from the analysis. However, a trial with negative results per se, if the regimen investigated had positive outcome in other trials, was not an exclusion criterion. In addition, trials including rectal cancer without a predefined colon cancer subgroup, were excluded from the analysis since radiotherapy is instrumental in (neo) adjuvant rectal cancer treatment which would make it difficult to assess the impact of chemotherapy. Meta-analyses that were not scoreable by the ESMO-MCBS scale were excluded for analysis as well (Consolidated Standards of Reporting Trials (CONSORT) diagram figure 1).

All studies meeting the inclusion and exclusion criteria were graded using form 1 of the ESMO-MCBS V1.1 based on OS or DFS results. Additionally, for non-inferiority trials, the grading was influenced by toxicity, quality of A and B representing a substantial level of clinical benefit.12 15 Currently, form 1 does not apply penalties for toxicity. However, patient advocates in consultation with the ESMO-MCBS working group have suggested that toxicity annotations should be introduced for treatments with a high prevalence of strong acute toxicity (AT) or LT and this is currently under consideration.

The validity of the ESMO-MCBS is predicated on adherence to the public policy ethical standard of ‘accountability for reasonableness’.12–15 Whereas the grading of treatments of advanced and incurable cancer using forms 2a, b, c and 3 of the ESMO-MCBS has been extensively field tested and reviewed for reasonableness, hitherto only 13 trials with curative intent, including adjuvant therapies, had been field tested using form 1. The main purpose was to evaluate the applicability of the scale in adjuvant colon cancer trials and further assess the reasonableness of the generated scores, in order to identify shortcomings that may be rectified in future amendments.

The European Society for Medical Oncology (ESMO) has developed a validated and reproducible tool to assess the magnitude of clinical benefit of anticancer therapies with either curative intent, such as adjuvant treatment in colon cancer or with non-curative intent. Form 1 has been developed for assessing new approaches to adjuvant therapy or new potentially curative therapies. The scale ranges from A to C, with grades
life (QoL) and costs. If there were up to three predefined subgroups included in the trial and there was an appropriate adjustment for multiplicity, these subgroups were graded individually. Trials that did not meet the criteria for scoring due to insufficient benefit (negative studies) have been designated as trials with ‘no evaluable benefit’ (NEB). Negative non-inferiority (NNI) studies were labelled as NNI. Extracted data and grading were reviewed by the ESMO-MCBS Working Group for accuracy.

An exploratory analysis of reported toxicity data was undertaken to determine the feasibility of toxicity annotations. Side effects during treatment or within 3 months after treatment completion were defined as AT. LT was defined as all events that occurred 3 months after treatment completion in accordance with Common Toxicity Criteria.28 AT as well as LT was annotated as overall less (+), equal (=) or more (+) toxicity for the intervention versus the control group. When there is insufficient data reported to draw conclusions, not reported (NR) is annotated.

The scores generated in this field testing were reviewed by the ESMO Gastro-Intestinal Tumours Faculty for reasonableness.

RESULTS

The literature search yielded 57 RCTs and 14 meta-analyses, with 18 RCTs and 7 meta-analyses finally found eligible and were included in the analysis. Reasons for exclusion for final analyses are summarised in the CONSORT diagram figure 1 and excluded studies and meta-analyses can be found in the supplementary references.

ESMO-MCBS grading

Information for the selected trials is summarised in table 1 for fluoropyrimidine regimens and in table 2 for oxaliplatin added to fluoropyrimidine regimens. Results in the tables are categorised to combined stage II and III, stage II and stage III.

Fluoropyrimidine regimens

Four trials and a meta-analysis compared 5-FU/LV chemotherapy with MOF combination chemotherapy (lomustine (MeCCNU), vincristine and non-modulated 5-FU)29 or surgery only for combined stage II and III colon cancers.30–33 They showed OS gain ranging from 5% to 14% at 3.0–5.0 years, resulting in the highest-grade ESMO-MCBS grade (A). These results were confirmed in three successive meta-analyses by the Adjuvant Colon Cancer End-points (ACCENT) Group showing a 7.0%–7.2% OS advantage at 5–8 years follow-up (grade A).24–26 Since 5-FU/levamisole (LEV) was included in these meta-analyses, the OS benefit of 5-FU/LV was probably underestimated since LEV was subsequently found to be inferior to LV as a 5-FU modulator.37–39 Table 1.

The two trials with 318 and 500 patients31,32 and a meta-analysis33 with 1016 patients evaluated adjuvant 5-FU/LV versus no adjuvant therapy in stage II colon cancer. None of these three studies demonstrated OS benefit and all were annotated as NEB. A 2004 meta-analysis involving 1440 patients35 demonstrated a non-significant 5-year survival gain of 1% (ESMO-MCBS grade NEB), however, a subsequent 2009 evaluation by the same group36 with more mature data reported a 5.4% OS benefit at 8 years (grade A). This discrepancy is addressed in the discussion below.

In grade III colon cancer, the observed OS benefit was 13.5% and 10.3% at 5 and 8 years, respectively, resulting in a grade A on the ESMO-MCBS.36,37

Uracil and tegafur (UFT)/LV in combined stage II and III colon cancer,38–40 capcitabine38 and S-139,40 in stage III colon cancer all did not provide an OS or DFS benefit compared with 5-FU/LV. Three of the four trials were non-inferiority trials. Although non-inferiority was proven, since neither QoL nor toxicity was improved; all studies were graded NEB.

Fluoropyrimidines with oxaliplatin combinations

Oxaliplatin added to 5-FU based regimens was evaluated in the MOSAIc41 and NSABP CO742 trials including stage II and III patients and the NO16968 trial confined to stage III patients.43 Greater clinical benefit (grade A) was observed in the trial confined to stage III compared with the other two trials which were graded B and NEB, respectively. The ACCENT group published a meta-analysis of these studies in 2016.44 Based on 5-year OS data their analysis demonstrated an insignificant 0.8% OS gain for stage II (NEB) and a 4.2% OS advantage for stage III colon cancer (grade B). Table 2.

In 2018 the International Duration Evaluation of Adjuvant (IDEA) consortium reported the planned combined analysis of 6 individual RCTs, with a non-inferiority design, comparing folinic acid/5-FU/oxaliplatin (FOLFOX) and capcitabine/oxaliplatin (CAPOX) for 3 vs 6 months.45 The 3-year DFS rate was very similar but non-inferiority was not proven for the intention to treat population resulting in a NNI. A preplanned subgroup analysis showed that 3 months CAPOX was non-inferior compared with 6 months. The 3 months treatment arm received a grade B based on non-inferiority in combination with less toxicity.46–48 T4 versus T1-3 and N2 versus N1 subgroups were prespecified however their combinations in subgroups and its interaction test was not significant, thus these subgroup analyses were post hoc and could not be graded.

In one meta-analysis, in stage III patients, capcitabine with or without oxaliplatin versus 5-FU/LV with or without oxaliplatin was examined. As no OS and DFS benefit was seen, and neither QoL nor toxicity was improved, it was graded NEB.49

Toxicity, QoL and cost

AT and LT reported in the included trials are summarised in the online supplementary table 1. All trials reported AT using several different approaches to toxicity evaluation: one trial did not use any grading system,50 five did NR the grading system used,51,25,26–28 five used the WHO toxicity scoring system,20,23,30,33,35 and seven the common terminology.
<table>
<thead>
<tr>
<th>Trial name, year first publication</th>
<th>Intervention versus control</th>
<th>N</th>
<th>Primary outcome</th>
<th>Median follow-up</th>
<th>DFS control group</th>
<th>DFS gain</th>
<th>DFS HR</th>
<th>DFS control group</th>
<th>OS gain</th>
<th>OS HR</th>
<th>Toxicity*</th>
<th>QoL</th>
<th>ESMO-MCBS V1.1</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fluoropyrimidine treatment versus surgery or no active adjuvant therapy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Combined stage II and III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSABPC-03 1993</td>
<td>5-FU/LV versus MOF</td>
<td>1041</td>
<td>DFS</td>
<td>3 years</td>
<td>9%</td>
<td>64% at 3 years</td>
<td>P=0.0004</td>
<td>77% at 3 years</td>
<td>7%</td>
<td>P=0.003</td>
<td>AT+LT NR</td>
<td>A</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Siena 1994</td>
<td>5-FU/LV versus surgery</td>
<td>239</td>
<td>DFS</td>
<td>4.5 years</td>
<td>15%</td>
<td>59% at 5 years</td>
<td>65% at 5 years</td>
<td>14%</td>
<td>AT+LT NR</td>
<td>A</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis IMPACT1995</td>
<td>5-FU/LV versus placebo</td>
<td>1493</td>
<td>5-FU/LV 3.3 yrsPlacebo 3.1 years</td>
<td>9%</td>
<td>62% at 3 years</td>
<td>0.67 (0.56–0.80)</td>
<td>78% at 3 years</td>
<td>5%</td>
<td>0.77 (0.62–0.96)</td>
<td>AT+LT NR</td>
<td>B</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCCTG 8746511997</td>
<td>5-FU/LV versus placebo</td>
<td>309</td>
<td>DFS</td>
<td>6 years</td>
<td>16%</td>
<td>0.58 at 5 years</td>
<td>0.63 at 5 years</td>
<td>11%</td>
<td>P=0.02</td>
<td>AT+LT NR</td>
<td>A</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIVO-SITAC 011998</td>
<td>5-FU/LV versus placebo</td>
<td>869</td>
<td>5-FU/LV 5.4 yrsPlacebo 5.3 years</td>
<td>12%</td>
<td>0.54 at 5 years</td>
<td>0.65 at 5 years</td>
<td>7%</td>
<td>AT+LT NR</td>
<td>A</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Sargent et al. 2001</td>
<td>5-FU/LV or 5-FU/LEV versus surgery alone</td>
<td>3351</td>
<td>OS</td>
<td>5.17–8.54 years</td>
<td>0.58 at 5 years</td>
<td>0.68 (0.60–0.76)</td>
<td>0.64 at 5 years</td>
<td>7%</td>
<td>0.76 (0.68–0.85)</td>
<td>A</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Gill et al. 2004</td>
<td>5-FU/LV or 5-FU/LEV versus surgery</td>
<td>3302</td>
<td>DFS and OS</td>
<td>5 years</td>
<td>12%</td>
<td>0.55 at 5 years</td>
<td>0.64 at 5 years</td>
<td>7%</td>
<td>0.74 (0.66–0.83)</td>
<td>A</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Sargent et al. 2009</td>
<td>5-FU/LV or 5-FU/LEV versus surgery</td>
<td>4922</td>
<td>DFS</td>
<td>8 years</td>
<td>NS (HR=0.61)</td>
<td>0.543 at 8 years</td>
<td>7.20%</td>
<td>HR=0.74</td>
<td>P ≤0.001</td>
<td>A</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stage II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intergroup 00351990</td>
<td>5-FU/LV versus surgery</td>
<td>318</td>
<td>OS</td>
<td>7 years</td>
<td>71% at 7 years</td>
<td>72% at 7 years</td>
<td>0%</td>
<td>AT+LT NR</td>
<td>NEB</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis IMPACT B2 1999</td>
<td>5-FU/LV versus placebo</td>
<td>1016</td>
<td>DFS</td>
<td>5.8 years</td>
<td>73% at 5 years</td>
<td>0.83 (0.68–1.01)</td>
<td>80% at 5 years</td>
<td>2%</td>
<td>0.86 (0.64–1.01)</td>
<td>AT+LT NR</td>
<td>NEB</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Gill et al. 2004</td>
<td>5-FU/LV or 5-FU/LEV versus surgery</td>
<td>1440</td>
<td>DFS and OS</td>
<td>–</td>
<td>4%</td>
<td>72% at 5 years</td>
<td>80% at 5 years</td>
<td>1%</td>
<td>NS p=0.11</td>
<td>NEB</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCSG Schipping et al. 2007</td>
<td>5-FU/LV versus surgery</td>
<td>500</td>
<td>OS</td>
<td>95.6m (~8 years)</td>
<td>0.80% at 7 years</td>
<td>0.95 (0.69–1.31)</td>
<td>76.6% at 7 years</td>
<td>1.60%</td>
<td>0.88 (0.61–1.27)</td>
<td>AT+LT NR</td>
<td>NEB</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Sargent et al. 2009</td>
<td>5-FU/LV or 5-FU/LEV versus surgery</td>
<td>8 years</td>
<td>66.8% at 8 years</td>
<td>5.40%</td>
<td>0.026</td>
<td>A</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stage III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intergroup 00351990</td>
<td>5-FU/LV versus surgery alone</td>
<td>619</td>
<td>OS</td>
<td>6.5 years</td>
<td>43.8% at 5 years</td>
<td>46.7% at 5 years</td>
<td>13.50%</td>
<td>P≤0.001</td>
<td>AT+LT7</td>
<td>A</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Trial name</th>
<th>Year first publication</th>
<th>Intervention versus control</th>
<th>N</th>
<th>Primary outcome</th>
<th>Median follow-up</th>
<th>DFS control group</th>
<th>DFS gain DFS HR</th>
<th>OS control group</th>
<th>OS gain OS HR</th>
<th>Toxicity*</th>
<th>QoL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCT01349962</td>
<td>2009</td>
<td>5-FU/LV versus 5-FU/LEV</td>
<td>68 at 5 years</td>
<td>63% at 5 years</td>
<td>8 years</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>A</td>
<td>26</td>
</tr>
<tr>
<td>NSABP C-041999</td>
<td>2009</td>
<td>5-FU/LEV versus 5-FU/LV</td>
<td>68 at 5 years</td>
<td>63% at 5 years</td>
<td>8 years</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>A</td>
<td>26</td>
</tr>
<tr>
<td>Intergroup 0089</td>
<td>2004</td>
<td>RPMI‡ versus 5-FU/LEV</td>
<td>68 at 5 years</td>
<td>63% at 5 years</td>
<td>8 years</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>A</td>
<td>26</td>
</tr>
<tr>
<td>Mayo Clinic§ versus 5-FU/LEV</td>
<td>2004</td>
<td>68 at 5 years</td>
<td>63% at 5 years</td>
<td>8 years</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>A</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Mayo Clinic+LEV versus 5-FU/LEV</td>
<td>2004</td>
<td>68 at 5 years</td>
<td>63% at 5 years</td>
<td>8 years</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>A</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Stage III adjCCA-012001</td>
<td>2004</td>
<td>5-FU/LV versus 5-FU/LEV</td>
<td>68 at 5 years</td>
<td>63% at 5 years</td>
<td>8 years</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>A</td>
<td>26</td>
</tr>
<tr>
<td>ACTS-CC 2014 non-inferiority trial</td>
<td>2015</td>
<td>S-1 versus UFT/LV</td>
<td>68 at 5 years</td>
<td>63% at 5 years</td>
<td>8 years</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>42.7% at 8 years</td>
<td>0.001</td>
<td>A</td>
<td>26</td>
</tr>
</tbody>
</table>

*Toxicity* and QoL data not available for all trials.
Reporting was more complete in the latest reported studies. The most recent IDEA consortium trial was the most complete in acute adverse events reporting, summarised in online supplementary table 1. This non-inferiority trial was the only trial in which AT data influenced the grade, as one prespecified subgroup with non-inferior efficacy was rewarded for having less AT to a B grade.

Reporting of LT was very limited. Five trials (two were individual trials within the IDEA collaboration), reported late sensory neuropathy graded with the CTCAE V.1–3.41–43 45–47 In all trials, this was investigator reported data and the assessment times and follow-up period differed. Overall, the reported prevalence of late neuropathy was low. With regard to oxaliplatin treatment duration, in the IDEA France trial, at a median follow-up of 3.6 years, the prevalence of grade 3–4 neuropathy was 0.5% among patients exposed to 3 months of oxaliplatin versus 2% among those who received 6 months of oxaliplatin.46

In the ACHIEVE trial, at a median follow-up of 3 years, the prevalence all grade neuropathy was 23.3% vs 10%, while grade 3 was only 0.3 vs 0%.47 In the ACHIEVE trial, it was also observed that the incidence of any grade PSN was lower for patients treated with CAPOX compared with FOLFOX in both the 6 months and 3 months treatment groups. All other studies did NR any LT.

QoL data were only available for 5 of the 20 RCT (one was an individual trial within the IDEA collaboration).23 35–38 48 There was no consistency in the scales used. The only trial to report differences in QoL between the treatment arms was the SCOT trial of the IDEA consortium which compared 3–6 months of oxaliplatin based adjuvant therapy. After 3 months to 5 years of follow-up, there was major difference (p<0.001) in neuropathy-related QoL evaluated using the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity questionnaire. Patients receiving 6 months oxaliplatin reported a worse QoL at 1, 3 and 5 years compared with those receiving 3 months oxaliplatin, and this disparity was associated with major differences in Global QoL between 3 and 6 months gradually attenuating over subsequent months and years.48

None of the trials did a formal cost analysis and could therefore not be used for grading of non-inferiority trials. However, non-inferiority of a shorter treatment duration most likely leads to reduction in treatment cost.

### Expert peer review of the generated scores

Thirty-seven experts from the ESMO Gastro-Intestinal Tumours Faculty reviewed the generated scores. Twenty-five (67.6%) confirmed the reasonableness of the scores and 12 (34.4%) found the scores mostly reasonable. Experts pointed out that it was striking that the current recommended oxaliplatin-based treatment for stage III disease was only once graded with the highest A grade.

Two experts commented on non-inferiority trials that offer a similar efficacy despite evaluating a more convenient oral mode of administration. They expressed

---

### Table 1

<table>
<thead>
<tr>
<th>Trial name, year first publication</th>
<th>Intervention versus control</th>
<th>Primary outcome</th>
<th>Median follow-up</th>
<th>OS control group</th>
<th>DFS control group</th>
<th>OS gain DFS</th>
<th>Control HR</th>
<th>Toxicity*</th>
<th>QoL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AT, acute toxicity; DFS, disease-free survival; EFS, event free survival; ESMO- European Society for Medical Oncology-Magnitude of Clinical Benefit Scale; 5-FU, 5-fluorouracil; HDLV, high dose leucovorin; HR, hazard ratio; LDLV, low dose leucovorin; LEV, levamisole; LT, late toxicity; LV, leucovorin; MOF, lomustine (MeCCNU)/vincristine/5-fluorouracil; N, number of patients; NEB, no evaluable benefit; NR, not reported; NS, not significant; OS, overall survival; QoL, quality of life; RPMI, Roswell Park Memorial Institute; S-1, tegafur/gimeracil/oteracil; UFT, uracil and tegafur; vs, versus.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*All toxicity annotations are exploratory; not part of the latest ESMO-MCBS forms. Toxicity of experimental arm versus control arm, more or less acute toxicity (duration of treatment and thus exposure time to drug and toxicity accounted for as well) of the experimental arm versus the control arm is shown as AT+ or LT+. More or less late toxicity is shown as LT+ or LT-. All toxicity data are summarised in online supplementary table 1. |

†The endpoint in the IMPACT B2 meta-analysis was EFS, however the given definition ‘time from randomisation to first event (ie, either a first recurrence, second tumour, or death from any cause)’ is not different from the up’. For better readability DFS is shown in the table instead of EFS. |

‡RPMI; 5-FU+HDLV for six courses. |

*Reporting of LT was very limited. Five trials (two were individual trials within the IDEA collaboration), reported late sensory neuropathy graded with the CTCAE V.1–3.41–43 45–47 In all trials, this was investigator reported data and the assessment times and follow-up period differed. Overall, the reported prevalence of late neuropathy was low. With regard to oxaliplatin treatment duration, in the IDEA France trial, at a median follow-up of 3.6 years, the prevalence of grade 3–4 neuropathy was 0.5% among patients exposed to 3 months of oxaliplatin versus 2% among those who received 6 months of oxaliplatin.46 In the ACHIEVE trial, at a median follow-up of 3 years, the prevalence all grade neuropathy was 23.3% vs 10%, while grade 3 was only 0.3 vs 0%.47 In the ACHIEVE trial, it was also observed that the incidence of any grade PSN was lower for patients treated with CAPOX compared with FOLFOX in both the 6 months and 3 months treatment groups. All other studies did NR any LT. |

QoL data were only available for 5 of the 20 RCT (one was an individual trial within the IDEA collaboration).23 35–38 48 There was no consistency in the scales used. The only trial to report differences in QoL between the treatment arms was the SCOT trial of the IDEA consortium which compared 3–6 months of oxaliplatin based adjuvant therapy. After 3 months to 5 years of follow-up, there was major difference (p<0.001) in neuropathy-related QoL evaluated using the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity questionnaire. Patients receiving 6 months oxaliplatin reported a worse QoL at 1, 3 and 5 years compared with those receiving 3 months oxaliplatin, and this disparity was associated with major differences in Global QoL between 3 and 6 months gradually attenuating over subsequent months and years.48 None of the trials did a formal cost analysis and could therefore not be used for grading of non-inferiority trials. However, non-inferiority of a shorter treatment duration most likely leads to reduction in treatment cost. |
### Table 2  
ESMO-MCBS grades of oxaliplatin added to fluoropyrimidine treatment for combined stage II and III, stage III and stage II colon cancer

<table>
<thead>
<tr>
<th>Trial name, year first publication</th>
<th>Intervention versus control</th>
<th>N</th>
<th>Primary outcome</th>
<th>Median follow-up</th>
<th>DFS control group</th>
<th>DFS gain</th>
<th>DFS HR</th>
<th>OS control group</th>
<th>OS Gain</th>
<th>OS HR</th>
<th>Toxicity*</th>
<th>QoL</th>
<th>ESMO-MCBS V1.1</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage II and II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOSAIC 2004</td>
<td>FOLFOX4 versus 5-FU/LV</td>
<td>2246</td>
<td>DFS</td>
<td>9.5 yrs</td>
<td>61.7% at 10 yrs</td>
<td>5.80%</td>
<td>0.82 (0.71–0.95)</td>
<td>67.1% at 10 yrs</td>
<td>4.60%</td>
<td>0.85 (0.73–0.99)</td>
<td>AT+ LT+</td>
<td>B</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>NSABP C-07 2007</td>
<td>FLOX versus 5-FU/LV</td>
<td>2409</td>
<td>DFS</td>
<td>8 yrs</td>
<td>64.2% at 5 yrs</td>
<td>5.20%</td>
<td>0.82 (0.72–0.93)</td>
<td>78.4% at 5 yrs</td>
<td>1.80%</td>
<td>0.88 (0.75–1.02)</td>
<td>AT+ LT+</td>
<td>NEB</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Shah et al 2016 44</td>
<td>5-FU/LV + OX versus 5-FU/LV</td>
<td>6468</td>
<td>OS</td>
<td>6 yrs</td>
<td>77.7% at 5 yrs</td>
<td>2.30%</td>
<td>Significant (HR not presented)</td>
<td>77.7% at 5 yrs</td>
<td>2.30%</td>
<td>Significant (HR not presented)</td>
<td>C</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stage II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Shah et al 2016 44</td>
<td>5-FU/LV + OX versus 5-FU/LV</td>
<td>1600</td>
<td>OS</td>
<td>6 yrs</td>
<td>89.8% at 5 yrs</td>
<td>0.80%</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stage III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO16968 2011</td>
<td>CAPOX versus 5-FU/LV</td>
<td>1886</td>
<td>DFS</td>
<td>7 yrs</td>
<td>56% at 7 yrs</td>
<td>7%</td>
<td>0.8 (0.69–0.93)</td>
<td>67% at 7 yrs</td>
<td>6%</td>
<td>0.83 (0.70–0.99)</td>
<td>AT+ LT+</td>
<td>A</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Shah et al 2016 44</td>
<td>5-FU/LV + OX versus 5-FU/LV</td>
<td>4868</td>
<td>OS</td>
<td>6 yrs</td>
<td>73.7% at 5 yrs</td>
<td>4.20%</td>
<td>Significant (HR not provided)</td>
<td>73.7% at 5 yrs</td>
<td>4.20%</td>
<td>Significant (HR not provided)</td>
<td>B</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Duration of therapy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDEA consortium NCT00958737 2018</td>
<td>CAPOX/FOLFOX 3 versus 6 m</td>
<td>12834</td>
<td>DFS</td>
<td>(margin 1.12 for DFS)</td>
<td>41.8 m (-3.5 yrs)</td>
<td>75.5% at 3 yrs</td>
<td>-0.90%</td>
<td>1.07 (1.00–1.15)</td>
<td>AT- LT†</td>
<td>Overall =‡ NSN -‡</td>
<td>NNI</td>
<td>45–48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAPOX 3 versus 6 m</td>
<td></td>
<td>5071</td>
<td>DFS</td>
<td></td>
<td>74.8% at 3 yrs</td>
<td>1.10%</td>
<td>0.95 (0.85–1.06)</td>
<td>AT- LT†</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOLFOX 3 versus 6 m</td>
<td></td>
<td>7763</td>
<td>DFS</td>
<td></td>
<td>76% at 3 yrs</td>
<td>-2.40%</td>
<td>1.16 (1.06–1.26)</td>
<td>AT- LT-</td>
<td>NNI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Convenience of therapy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis Schmoll et al 2014</td>
<td>CAP + OX versus 5-FU/LV + OX</td>
<td>5819</td>
<td>DFS</td>
<td></td>
<td>CAP + OX 62.8% at 5 yrs</td>
<td>0%</td>
<td>1.01 (0.92–1.10)</td>
<td>73.9% at 5 yrs</td>
<td>-1.3%</td>
<td>1.02 (0.92–1.14)</td>
<td>AT = LT=</td>
<td>NEB</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>

*All toxicity annotations are exploratory; not part of the latest ESMO-MCBS forms. Toxicity of experimental arm versus control arm, more or less acute toxicity (duration of treatment and thus exposure time to drug and toxicity accounted for as well) of the experimental arm versus the control arm is shown as AT+ or AT–, more or less late toxicity is shown as LT+ or LT–. All toxicity data are summarised in Supplementary Table 4. Late toxicity data were not reported in the IDEA consortium pooled analyses. However, subsequently two of the six individual trials reported late toxicity data which is shown here.

†QoL data were not reported in the IDEA consortium pooled analyses. However, subsequently one of six individual trials reported QoL data which is reported here.

‡QoL, quality of life; yrs, years. **
concern that the failure to reward this difference in convenience may fail to credit a true benefit. The ESMO-MCBS V.1.1 of form I does not offer the means to credit convenience.

DISCUSSION

This paper has evaluated the applicability of form I of the ESMO-MCBS V.1.1 to the adjuvant therapies for early colon cancer. Overall, the experience has been positive insofar as the scoring of adjuvant approaches in early colon cancer are considered reasonable (67.6%) or mostly reasonable (32.4%) by all experts.

For patients with stage III colon cancer, RCTs and meta-analyses of modulated 5-FU-based chemotherapy versus surgery only, consistently scored A in the ESMO-MCBS and the RCTs and meta-analysis comprising oxaliplatin added to this 5-FU backbone showed a modest additional OS benefit (grade B).

For stage II, the findings were less consistent. Whereas fluoropyrimidine trials in patients with stage II colon cancer consistently were graded NBI, the most recent meta-analysis demonstrated a 5.4% survival advantage after 8 years follow-up (grade A). The ACCENT investigators have subsequently cautioned that conclusions derived from older trials of FU-based adjuvant therapy in stage II colon cancer may be biased by stage migration over time. To date, there are no subgroup analyses restricted to stage II in trials with patients that were adequately staged by contemporary standards. RCTs and meta-analysis adding oxaliplatin demonstrated no added benefit for patients with stage II colon cancer.

Several meta-analyses analysed efficacy in stage II/III, as well as separately in II or III, or stage II/III. Four of these were performed by the ACCENT Collaborative Group, which, as of 2016, included detailed information collected from over 40,000 patients from 27 adjuvant colon cancer trials including patient demographics and disease characteristics, treatment data, biomarkers for selected studies, adverse events, as well as long term recurrence and survival follow-up for all patients. This has facilitated the capacity to undertake robust analysis of pooled individual patient data in meta-analyses and in the evaluation of the validity of surrogate outcomes.

Regarding the surrogacy of DFS as a predictor of OS, analysis by the ACCENT Collaborative Group demonstrated a robust relationship for 2, 3, 5 and 6 years DFS and OS for stage III colon cancer, but this was not the case for stage II disease and indeed even 6 years DFS was only weakly associated with OS. Consequently, they concluded that unless DFS is considered a clinically relevant endpoint, OS should be regarded as the most appropriate endpoint for trials in unselected stage II disease.

The ESMO-MCBS V.1.1 has no defined rules regarding the minimum quality requisites for a meta-analysis to be evaluated. In future amendments of the scale, formal definitions of quality and improved clarity regarding the issue of multiplicity when there are several subgroup analyses will be important. In general, an impactful and valid meta-analysis should include at least the following ingredients: investigation of a plausible question based on randomised evidence using an exhaustive review of relevant studies; evaluation of consistency across studies regarding population of interest, relevant patient characteristics and control arm, coupled with lack of bias (publication, selective reporting); exploration of heterogeneity and clear description of limitations.

Reporting of toxicity and QoL effects of new adjuvant systemic treatment modalities, especially if long-lasting, is important to optimally inform patients. A penalty system for toxicities, such as used in the non-curative setting in the ESMO-MCBS V.1.1 (forms 2 and 3), is not appropriate for the curative setting (form I) since patients may accept higher toxicity trade-off when treatment is with curative intent. Representatives of patient advocacy groups, in consultation with the ESMO-MCBS Working Group, have indicated preference for annotation of high likelihood of AT or LT versus penalties which may mask the magnitude of curative potential. We strongly believe toxicity annotations should indeed be introduced for treatments with a high prevalence of AT and especially LT.

Our exploratory evaluation of toxicity highlighted that toxicity evaluation and annotation is challenging in the setting of inconsistent methods of toxicity reporting, a high prevalence of apparent under reporting and minimum reporting of LT. The chronic neurotoxicity induced by oxaliplatin is a cumulative, dose-dependent, sensory, symmetric distal axonal neuropathy. Tinning is the most prominent symptom, but numbness and pain can also occur. In our review of the toxicity data, late grade 3/4 PSN was reported in only 0.5%–2% of patients, substantially lower than the prevalence data derived from patient reported outcome data. This highlights the risk of under-reporting of toxicities by physicians. In addition, even several years after adjuvant oxaliplatin-based chemotherapy, in some situations distal neurotoxicity symptoms are reported as re-induced by cold temperature or repeated use of fingers like key-board typing, piano playing or exercising precise finger movements. This is general not mentioned in the toxicity report of clinical trial but has a potential negative impact on QoL or professional career.

In our analyses, only 5 out of 18 trials evaluated QoL. The low rate of inclusion of QoL evaluation has been examined in a study comprised by phase III RCTs in cancer performed between 2012 and 2016 published in 11 major journals. In 210 of the 446 trials (47.1%), QoL was not included as an endpoint. The non-inclusion was even higher for RCTs in (neo)adjuvant disease as 81 of the 124 trials (65.3%) did not include QoL as an endpoint. Most of the adjuvant trials reporting QoL showed no difference between the investigational and control arm: 5-FU/LV or placebo, UFT/LV or 5-FU/LV and capecitabine or 5-FU/LV. The findings of the SCOT trial demonstrated worse QoL for PSN...
at 1, 3 and 5 years for patients treated 6 vs 3 months were salient (p<0.001).

UFT/LV did not show OS benefit and a non-inferior OS for stage II/III colon cancer compared with 5-FU/LV in two trials and neither QoL nor toxicity was improved. Both trials were graded NEB as was the trial of capecitabine versus 5-FU/LV. While it is plausible that oral medication may be more convenient than intravenous treatment, there are no data that it actually improves QoL compared with conventional parenteral administration. Convenience is not credited in the current version of form 1 of the ESMO-MCBS.

Our findings confirm that form 1 was highly applicable to the studies of adjuvant systemic therapies of early-stage colon cancer and it provided very reasonable grading for adjuvant colon cancer studies. The exploratory analysis indicated that toxicity annotation is feasible but the prevailing convention of physician reported toxicity may underestimate the true level of patient burden from both AT and LT. Since patients in the curative setting potentially live decades after treatment, late and prolonged adverse effects that may undermine QoL should be annotated to optimally inform patients of recognised risks. Future revisions of form 1 of the ESMO-MCBS will be cognizant of these findings.

Author affiliations
1 Medical Oncology, University Medical Centre Groningen, Groningen, Groningen, Netherlands
2 Medical Oncology, Shaare Zedek Medical Center, Jerusalem, Jerusalem, Israel
3 Statistics, Frontier Science Foundation-Hellas, Statistics, Athens, Zografou, Greece
4 ESMO-MCBS Working Group, European Society for Medical Oncology, Viganello, Switzerland
5 Nursing, National and Kapodistrian University of Athens, Goudi-Athens, Greece
6 University of Athens, Athens, Greece

Acknowledgements
The authors wish to thank and acknowledge our oncology colleagues who participated in the field testing for ‘reasonableness’ of scorings using form 1 of the ESMO-MCBS and who have agreed to place their names in this publication (online supplementary table 2). We also thank those who wished to remain anonymous.

Contributors
Conception of the work.
Funding
ESMO (no grant numbers apply).

Competing interests
EEGdV reports Institutional Financial Support for her advisory role from Daiichi Sankyo, Merck, NSABP, Pfizer, Sanofi, Synthion and Institutional Financial Support for clinical trials or contracted research from Amgen, AstraZeneca, Bayer, Chugai Pharma, CytomX Therapeutics, Genentech, Nordic Nanovector, Radius Health, Regeneron, Roche, Synthon, all outside the submitted work.

Patient consent for publication
Not required.

Provenance and peer review
Commissioned; internally peer reviewed.

Data availability statement
All data relevant to the study are included in the article or uploaded as ONLINE supplementary information.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, any changes made are indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Panagiota Zygoura http://orcid.org/0000-0002-4867-8676
Derk Jan de Groot http://orcid.org/0000-0002-8306-6835

REFERENCES


Open access
First published as 10.1136/esmoopen-2020-000681 on 6 September 2020. Downloaded from esmoopen.bmj.com on September 16, 2020 by University of Groningen.

Open access


53 Sargent D, Shi Q, Yothers G, et al. Two or three year disease-free survival (DFS) as a primary end-point in stage III adjuvant colon cancer trials with fluoropyrimidines with or without oxaliplatin or irinotecan: data from 12,676 patients from mosaic, XE-301, N0182, PETACC-3, C-06, C-07 and C89803. J Clin Oncol 2014;32:195–200.


