Evaluation and analysis of stepped wedge designs
Zhan, Zhuozhao

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
5

SURVIVAL ANALYSIS OF THE CEAwatch MULTICENTRE CLUSTERED RANDOMIZED TRIAL

C. J. Verberne
Z. Zhan
E. R. van den Heuvel
F. Oppers
A. M. de Jong
I. Grossmann
J. M. Klaase
G. H. de Bock
T. Wiggers

This chapter has been published in British Journal of Surgery, Volume 104, Issue 8, (2017) [20]. (C.J.V. and Z.Z. contributed equally to this work)
ABSTRACT

**Background:** The CEAwatch randomized trial showed that follow-up with intensive carcinoembryonic antigen (CEA) monitoring (CEAwatch protocol) was better than care as usual (CAU) for early postoperative detection of colorectal cancer recurrence. The aim of this study was to calculate overall survival (OS) and disease-specific survival (DSS).

**Methods:** For all patients with recurrence, OS and DSS were compared between patients detected by the CEAwatch protocol versus CAU, and by the method of detection of recurrence, using Cox regression models.

**Results:** Some 238 patients with recurrence were analysed (7.5 per cent); a total of 108 recurrences were detected by CEA blood test, 64 (55.2 per cent) within the CEAwatch protocol and 44 (41.9 per cent) in the CAU group (P = 0.007). Only 16 recurrences (13.8 per cent) were detected by patient self-report in the CEAwatch group, compared with 33 (31.4 per cent) in the CAU group. There was no significant improvement in either OS or DSS with the CEAwatch protocol compared with CAU: hazard ratio 0.73 (95 per cent 0.46 to 1.17) and 0.78 (0.48 to 1.28) respectively. There were no differences in survival when recurrence was detected by CT versus CEA measurement, but both of these methods yielded better survival outcomes than detection by patient self-report.

**Conclusion:** There was no direct survival benefit in favour of the intensive programme, but the CEAwatch protocol led to a higher proportion of recurrences being detected by CEA-based blood test and reduced the number detected by patient self-report. This is important because detection of recurrence by blood test was associated with significantly better survival than patient self-report, indirectly supporting use of the CEAwatch protocol.
5.1. INTRODUCTION

Patients with colorectal cancer who have undergone surgery with curative intent usually require follow-up because early detection of asymptomatic disease improves the probability of treatment success. Indeed, the options for cure in recurrent disease have increased over time, leading to higher cure rates in this patient group. [4] Although postoperative surveillance guidelines after colorectal cancer surgery differ between countries, there is consensus that imaging and carcinoembryonic antigen (CEA) measurements should be performed routinely. Current Dutch guidelines on follow-up recommend measurement of CEA levels, along with liver ultrasonography and chest X-rays every 3–6 months for the first 3 years (http://www.oncoline.nl/coloreectaalcarcinoom).

Recent research has focused on whether intensifying current follow-up strategies could improve survival, and several studies have compared the survival benefits between different follow-up strategies. [13] Earlier studies showed a modest and clinically relevant gain in survival for intensive protocols [10], but more recent trials [14, 15] have not reported such benefits. As available treatments and imaging technologies for the detection of recurrent disease have improved, data from older studies are likely to be invalid in the current era.

The CEAwatch RCT [18] showed that intensive CEA-based postoperative screening after colorectal cancer surgery (CEAwatch protocol) could provide benefits in terms of earlier detection times and higher curative treatment rates compared with care as usual (CAU) following current Dutch guidelines. In this trial, the CEAwatch protocol consisted of CEA measurements every 2 months in the first 3 years, and once every 3 months during the fourth and fifth years, combined with annual imaging by CT.

The aim of the present analysis was to assess whether the shortened
5. CEA-Watch: Survival outcomes

detection time and increased curative treatment rate have been associated with an increase in overall (OS) and disease-specific (DSS) survival. A secondary objective was to investigate whether there has been a difference in survival in relation to the method of detection of recurrence, namely CT imaging, CEA-based blood test and patient self-report. The null hypothesis of the present study was that patients with recurrence detected by the CEAwatch protocol would have the same risk of OS and DSS as those with recurrence detected by the CAU protocol.

5.2. Methods

The CEAwatch trial was a multicentre, stepped-wedge, cluster randomized trial in the Netherlands, which compared an intensive CEA-based follow-up protocol with the current national guideline (CAU). The trial had a unidirectional crossover design in which each cluster switched from control treatment to intervention treatment, with randomization used to allocate clusters to different switch moments. Eleven non-academic teaching hospitals were randomly grouped into five clusters. The trial started in October 2010, and every 3 months one cluster switched from the CAU follow-up protocol to the intensified follow-up protocol (Fig. 5.1). A detailed description of the trial design has been published elsewhere.[22] The CEAwatch trial (Netherlands Trial Register NTR2182) was approved by the Medical Ethics Committee of the University Medical Centre Groningen (METc-UMCG 2010.064) and by the local ethics committees of all participating centres.

5.2.1. Participants and data collection

Patients were eligible for inclusion if they had primary colorectal cancer, AJCC stage I–III disease, and underwent R0 resection between 2007 and July 2012. Between October 2010 and July 2012, all patients who provided
5.2. Methods

Figure 5.1 | Flow chart for patients with colorectal carcinoma in the CEAwatch trial. *Present in the Dutch Surgical Colorectal Audit (DSCA) database, but not found in hospital database

Informed consent were assigned to the protocol of the hospital they were attending, and followed up until March 2015. For the present analysis, data on survival status (alive, dead), oncological status (recurrence, no recurrence), cause of death (cancer-related, not cancer-related), methods of detection of recurrence, and the treatment employed for recurrences were collected. The method of detection of recurrence was defined as that which indicated an abnormality (CEA, physical signs) leading to the imaging confirming recurrent disease. Data were updated until March
2015 by three investigators.

5.2.2. Follow-up protocols

The control (CAU) follow-up protocol followed the national guidelines of the Netherlands in 2008. This comprised outpatient clinic visits every 6 months for the first 3 years and annual visits in years 4 and 5. CEA measurement was recommended every 3–6 months in the first 3 years and annually in following 2 years. It is known that adherence to this guideline is poor regarding the use of CEA measurement.[6] Liver ultrasonography and chest X-ray were recommended at each visit. Postoperative follow-up started after curative resection and adjuvant therapy, if this was given. Dutch oncology guidelines advise discussion of adjuvant chemotherapy with the patient in the case of AJCC stage III and II colonic cancer with a high risk of recurrence (small numbers of retrieved lymph nodes, perforated tumours, T4 N0 and other poor prognostic characteristics). Adjuvant treatment for rectal cancer is not advised in the Netherlands.

The intensive (CEAwatch) protocol consisted of taking CEA measurements every 2 months and performing annual imaging with high resolution CT of the thorax and abdomen for the first 3 years, followed by CEA measurement every 3 months in the fourth and fifth years. In this protocol, if the absolute CEA value was greater than 2.5 ng/ml and there was a 20 per cent increase in the measured CEA value from the previous one, another blood sample was drawn 4 weeks later.[7] If a rise in CEA level was noted between consecutive measurements, CT of the chest and abdomen was advised (Fig. 5.2). This intensive CEA protocol was based on a single-centre phase II trial, in which CEA levels were checked every month.[7] Based on these results, the authors were able to define the schedule with the highest sensitivity and specificity for detection of recurrence. Overall, the intensive CEAwatch protocol has more frequent CEA measurements and fewer clinical visits than the CAU protocol.
Figure 5.2 | Schematic illustration of the intensive carcinoembryonic antigen (CEA) measurement (CEAwatch) and control (care as usual) follow-up protocols. *Local differences and adjustment by individual hospitals allowed.

### 5.2.3. STATISTICAL ANALYSIS

Demographic and clinical characteristics of groups with and without recurrence during the study period were compared. Baseline data and recurrence detection methods were compared between the two follow-up protocols among patients with recurrence. The Mann–Whitney $U$ test was used for analysis of continuous variables, and the $\chi^2$ test for categorical variables or Fisher’s exact test in case of insufficient cell count.

To investigate the effect of different follow-up strategies on survival, the analyses focused on patients with recurrence detected during the trial as the follow-up protocol is unlikely to influence survival among healthy (cured) patients. For both OS and DSS, a Cox Markov model [1, 12] was
used to compare the transition from recurrence to death among those with recurrence detected by the CEAwatch protocol versus the CAU protocol. The model was adjusted for age at diagnosis, sex, primary tumour stage and hospital. Hazard ratios (HRs) and 95 per cent confidence intervals were calculated.

Proportionality assumptions were verified by checking interactions between detection methods and logarithm of survival time, and by examining Schoenfeld residuals.[5, 17] When comparing the different detection methods, similar models were fitted by adjusting for the same co-variables and taking into account interactions between the detection methods and the follow-up protocol. If no significant effect was observed for the interaction, a model without these interactions was used in the final analysis.

For OS, survival curves were plotted for comparison both of the follow-up protocols and the detection methods. The survival functions were calculated by the Breslow estimator [3], taking into account the other co-variables instead of using the typical product-limit estimator of Kaplan–Meier curves. Adjusted survival curves are presented for the modal patient (typical patient with the most common characteristics) in the cohort.

For both OS and DSS, additional sensitivity analyses were performed to check the Markovian assumptions, that is whether the probability of death depended on the time of detection of a recurrence. For this, likelihood ratio tests were performed by including the time to detection as a co-variable in the Cox proportional hazard model.[12] In addition, an assessment was made of the assumption that the follow-up protocol had no effect on survival outcomes for patients without recurrence. This was done by fitting a Cox regression model to the data from the group without recurrence with follow-up protocol as a time-dependent co-variable. Two-sided P values are reported for all analyses; P <0.050 was considered significant. All statistical analyses were conducted using SAS® 9.4
statistical software (SAS Institute, Cary, North Carolina, USA).

5.3. RESULTS

Between October 2010 and July 2012, a total of 3223 patients were included in the CEAwatch trial. Forty-one patients were excluded (based on inclusion/exclusion criteria during data analysis, 29; missing data on recurrence, 3; secondary tumour rather than recurrence, 9), leaving 3182 patients in the final analysis. A comparison of the baseline characteristics between patients with and without recurrence is shown in Table 5.1.

<table>
<thead>
<tr>
<th>Table 5.1</th>
<th>Baseline characteristics of patients and cancers by recurrence within the trial period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recurrence in trial period</td>
</tr>
<tr>
<td></td>
<td>No (n = 2944)</td>
</tr>
<tr>
<td>Age at diagnosis (years)*</td>
<td>70 (63–77)</td>
</tr>
<tr>
<td>Sex ratio (M:F)</td>
<td>1630:1314</td>
</tr>
<tr>
<td>AJCC tumour stage</td>
<td></td>
</tr>
<tr>
<td>I–II</td>
<td>2016 (68.5)</td>
</tr>
<tr>
<td>III</td>
<td>928 (31.5)</td>
</tr>
<tr>
<td>Primry tumour location</td>
<td></td>
</tr>
<tr>
<td>Colon (+ rectosigmoid)</td>
<td>1858 (63.1)</td>
</tr>
<tr>
<td>Rectum</td>
<td>1086 (36.9)</td>
</tr>
<tr>
<td>Adjuvant chemotherapy</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2204 (74.9)</td>
</tr>
<tr>
<td>Yes</td>
<td>740 (25.1)</td>
</tr>
<tr>
<td>Patients with co-morbidity</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>566 (45.1)</td>
</tr>
<tr>
<td>Minor</td>
<td>593 (47.2)</td>
</tr>
<tr>
<td>Major</td>
<td>97 (7.7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1688</td>
</tr>
</tbody>
</table>

* Values in parentheses are percentages unless indicated otherwise; values are median (i.q.r.).
† χ² test, except
‡ Mann–Whitney U test and
§ Fisher’s exact test.

In total, 238 patients (7.5 per cent) had recurrent disease detected
during the trial period; the median age at diagnosis of the primary tumour was 69 years. Of these, 153 (64.3 per cent) were men, 128 (53.8 per cent) had an AJCC stage III tumour, 145 (60.9 per cent) had primary tumours in the colon (including rectosigmoid), and 80 (33.6 per cent) underwent adjuvant chemotherapy. The baseline characteristics of patients with recurrence detected by the CEAwatch and CAU follow-up protocols are compared in Table 5.2. Apart from a significant difference in age at diagnosis, the main difference between the two groups was in the method by which recurrence was detected. The proportion of patients with recurrence detected by imaging was similar for both protocols, but a significantly higher proportion had recurrence detected by a CEA-based blood test rather than patient self-report in the CEAwatch compared with the CAU group.

5.3.1. OVERALL SURVIVAL

There was no significant difference in OS between patients with recurrence diagnosed within the CEAwatch protocol and those whose recurrence was detected with CAU (HR 0.73, 95 per cent c.i. 0.46 to 1.17) (Table 5.3 and Fig. 5.3). Patients who were older at diagnosis had a significantly higher risk of death than younger patients (HR 1.02, 1.00 to 1.04) and those with AJCC stage III disease had a higher risk of death than those with stage I or II (HR 1.48, 1.03 to 2.14).

For recurrences detected by the same method (CEA or imaging), there were no differences in survival between the two follow-up protocols (P = 0.496 for the overall interaction between detection method and follow-up protocol). There were also no statistically significant differences in the risk of death between patients whose recurrences were detected by CEA-based blood test and those detected by imaging (HR 1.34, 0.83 to 2.17) (Table 5.4 and Fig. 5.4). However, both CEA measurement (HR 0.39, 0.25 to 0.63) and imaging (HR 0.29, 0.17 to 0.51) were associated with a significantly lower
Table 5.2 | Baseline characteristics of patients with recurrence according to follow-up protocol

<table>
<thead>
<tr>
<th>Follow-up protocol</th>
<th>p‡¶</th>
<th>Care as usual (n = 112)</th>
<th>CEAwatch (n = 126)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis (years)†</td>
<td>74 (64–80)  </td>
<td>66 (61–74)  </td>
<td>&lt;0.001#</td>
</tr>
<tr>
<td>Sex ratio (M:F)</td>
<td>68:44</td>
<td>85:41</td>
<td>0.283</td>
</tr>
<tr>
<td>AJCC tumour stage</td>
<td>0.842</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I–II</td>
<td>51 (45.5)</td>
<td>59 (46.8)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>61 (54.5)</td>
<td>67 (53.2)</td>
<td></td>
</tr>
<tr>
<td>Primry tumour location</td>
<td>0.839</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon (+ rectosigmoid)</td>
<td>69 (61.6)</td>
<td>76 (60.3)</td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>43 (38.4)</td>
<td>50 (39.7)</td>
<td></td>
</tr>
<tr>
<td>Detection method†</td>
<td></td>
<td></td>
<td>0.007**</td>
</tr>
<tr>
<td>Blood test (CEA)</td>
<td>44 (41.9)‡</td>
<td>64 (55.2)</td>
<td></td>
</tr>
<tr>
<td>Imaging</td>
<td>28 (26.7)</td>
<td>36 (31.0)</td>
<td></td>
</tr>
<tr>
<td>Patient self-report</td>
<td>33 (31.4)</td>
<td>16 (13.8)</td>
<td></td>
</tr>
</tbody>
</table>

Values in parentheses are percentages unless indicated otherwise;
* values are median (i.q.r.).
‡ Information missing for 17 patients.
‡ Three recurrences detected by both carcinoembryonic antigen (CEA) level and patient self-report, and one by both CEA level and imaging.
§ One recurrence detected by both CEA level and patient self-report.
¶ χ² test, except
# Mann–Whitney U test and
** Fisher’s exact test.
Table 5.3 | Cox Markov model analysis to determine the effect of follow-up protocols on overall survival adjusted for patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>Hazard ratio</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (per year)</td>
<td>1.02 (1.00, 1.04)</td>
<td>0.026</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.81 (0.56, 1.17)</td>
<td>0.261</td>
</tr>
<tr>
<td>F</td>
<td>1.00 (reference)</td>
<td></td>
</tr>
<tr>
<td>AJCC tumour stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>1.48 (1.03, 2.14)</td>
<td>0.035</td>
</tr>
<tr>
<td>I–II</td>
<td>1.00 (reference)</td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEAwatch</td>
<td>0.73 (0.46, 1.17)</td>
<td>0.191</td>
</tr>
<tr>
<td>Care as usual</td>
<td>1.00 (reference)</td>
<td></td>
</tr>
</tbody>
</table>

Values in parentheses are 95 per cent confidence intervals. The multivariable model was also adjusted for hospital. Check of dependency on time of detection of recurrence: parameter estimate -0.0005 (s.e. 0.0004); likelihood ratio 1.67, 1 d.f., P = 0.197.

risk of death than patient self-report (Table 5.4).

Table 5.4 | Comparison of the hazard ratio for overall survival between the different screening methods

<table>
<thead>
<tr>
<th></th>
<th>Reference group</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEA-based blood test</td>
<td>Imaging</td>
<td>1.34 (0.83, 2.17)</td>
</tr>
<tr>
<td>Imaging</td>
<td>Patient self-report</td>
<td>0.29 (0.17, 0.51)</td>
</tr>
<tr>
<td>CEA-based blood test</td>
<td>Patient self-report</td>
<td>0.39 (0.25, 0.63)</td>
</tr>
</tbody>
</table>

Values in parentheses are 95 per cent confidence intervals. CEA, carcinoembryonic antigen.

5.3.2. Disease-specific survival

The follow-up protocol (CEAwatch versus CAU) had no impact on the risk of death from colorectal cancer (HR 0.78, 95 per cent c.i. 0.48 to 1.28). However, older patients had a significantly higher risk of disease-
5.3. RESULTS

Figure 5.3 | Comparison of predicted overall survival for a typical modal patient with recurrence in two follow-up groups: intensive carcinoembryonic antigen measurement (CEAwatch) and care as usual (CAU)

specific death (HR 1.02, 1.00 to 1.04), as did patients with a higher tumour stage (AJCC III versus I–II) (HR 1.70, 1.14 to 2.52) (Table 5.5). As regards detection method, no differences in risk were observed between the CEA-based blood test and imaging methods (Table 5.6). In contrast, recurrence detected by CEA-based blood test (HR 0.33, 0.20 to 0.55) or imaging (HR of 0.26, 0.14 to 0.47) had a lower risk of colorectal cancer death than patient self-report.

5.3.3. SENSITIVITY ANALYSES

In verifying the assumption that the OS time did not differ between the two follow-up groups, the sensitivity analyses produced a non-significant HR of 0.75 (95 per cent c.i. 0.52 to 1.07), supporting the assumptions made in the main analyses.
Table 5.5 | Cox Markov model analysis to determine the effect of follow-up protocols on disease-specific survival adjusted for patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>Hazard ratio</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (per year)</td>
<td>1.02 (1.00, 1.04)</td>
<td>0.044</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.78 (0.53, 1.16)</td>
<td>0.217</td>
</tr>
<tr>
<td>F</td>
<td>1.00 (reference)</td>
<td></td>
</tr>
<tr>
<td>AJCC tumour stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>1.70 (1.14, 2.52)</td>
<td>0.009</td>
</tr>
<tr>
<td>I–II</td>
<td>1.00 (reference)</td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEAwatch</td>
<td>0.78 (0.48, 1.28)</td>
<td>0.328</td>
</tr>
<tr>
<td>Care as usual</td>
<td>1.00 (reference)</td>
<td></td>
</tr>
</tbody>
</table>

Values in parentheses are 95 per cent confidence intervals. The multivariable model was also adjusted for hospital. Check of dependency on time of detection of recurrence: parameter estimate -0.0004 (s.e. 0.0004); likelihood ratio 1.29, 1 d.f., P = 0.255.

Table 5.6 | Comparison of the hazard ratio for disease-specific survival between the different screening methods

<table>
<thead>
<tr>
<th></th>
<th>Reference group</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEA-based blood test</td>
<td>Imaging</td>
<td>1.28 (0.76, 2.14)</td>
</tr>
<tr>
<td>Imaging</td>
<td>Patient self-report</td>
<td>0.26 (0.14, 0.47)</td>
</tr>
<tr>
<td>CEA-based blood test</td>
<td>Patient self-report</td>
<td>0.33 (0.20, 0.55)</td>
</tr>
</tbody>
</table>

Values in parentheses are 95 per cent confidence intervals. CEA, carcinoembryonic antigen.
5.4. DISCUSSION

In this study, the intensive CEA-based follow-up protocol of the CEAwatch trial did not improve the OS or DSS of patients with recurrence. Similarly, the way in which recurrent disease was detected had no influence on either OS or DSS in comparisons of intensive CEA monitoring, conventional CEA measurements and imaging. However, OS and DSS were significantly worse when recurrences were detected by patient self-report alone compared with intensive CEA measurements (CEAwatch), conventional CEA measurements or imaging.

The findings of this study are in line with recent literature, including a meta-analysis.[13] In the FACS (Follow-up After Colorectal Surgery) trial [14], an RCT of 1202 patients with curatively resected colorectal cancer, participants were assigned to one of four follow-up groups: CEA alone
(300 patients), CT alone (299), CEA and CT (302) or minimal follow-up (301). The CEA level was considered raised to a degree that justified further investigation if it crossed an absolute threshold value of 7 ng/ml. In contrast to the CEAwatch trial, a dynamic threshold was not used, and the FACS trial reported a mean recurrence rate per year of 3.8 per cent after a mean 4.4 years of observation. Compared with minimal follow-up, intensive imaging or CEA screening increased the rates of surgical treatment with curative intent in patients with recurrence. As in the CEAwatch trial, intensive follow-up had no significant survival benefits over minimal follow-up, and all follow-up methods were more effective than no follow-up because of the poorer survival associated with waiting for recurrence to become symptomatic. The authors of the FACS trial suggested that the lack of survival gain with more intensive follow-up may have been due to a lack of study power to detect differences, and they intend to report again after a further 5 years of follow-up.

Similar conclusions were reached in the GILDA (Gruppo Italiano di Lavoro per la Diagnosi Anticipata) trial [15], an RCT comparing imaging follow-up strategies for colorectal cancer, in which 1228 patients were randomized to either minimal or intensive surveillance. Liver ultrasonography was performed twice over the 5-year study period in the minimal surveillance group, compared with every 4 months in the first year and often combined with chest X-rays in the intensive surveillance group. This contrasts with the CEAwatch trial, in which imaging was performed only once a year in the first 3 years. The mean recurrence rate per year was 4 per cent after a median follow-up of 5.2 years, which was higher than the 2.5 per cent per year observed in the CEAwatch trial. Despite this difference in imaging frequency, comparison of OS curves for the whole population showed no statistically significant differences. Although the COLOFOL study – a pragmatic randomized study to assess the frequency of surveillance tests after curative resection in patients with stage II and
III colorectal cancer – is now closed, no definitive results have been published. In the most recent report [2], published in January 2016, it was concluded that patients selected for COLOFOL were representative of the patient population suitable for follow-up.

With consistent evidence available from these three recent large-scale RCTs, the results of older studies, which suggested a clear survival gain with intensive follow-up, should no longer be considered relevant to the discussion of appropriate follow-up. These older studies are limited by more recent advances in diagnosis and therapy. Indeed, not only were they carried out before the routine use of preoperative CT, which has led to an increase in detection of synchronous metastases [8], they were also undertaken before critical advances were made in neoadjuvant and adjuvant treatments, as well as improvements in surgical techniques that have improved outcomes.[9, 11]

It has become clear from the FACS, GILDA and CEAwatch trials that both imaging and CEA measurement lead to the earlier detection of colorectal cancer recurrence during follow-up, but that regardless of the follow-up programme (imaging, CEA, or both), this is not directly translated into clear improvements in survival. However, a protocol with no follow-up was not included in any of the trials; even in the FACS study, a single CT scan was requested at study entry in the minimal follow-up arm. Given that there is sufficient evidence that treatment outcomes are worse for symptomatic than asymptomatic metastases, studies comparing follow-up with no follow-up after colorectal cancer treatment are unlikely to be justifiable. The present analysis confirmed this by showing significantly worse survival outcomes associated with recurrence detected by patient self-report than by CEA measurement or imaging. The CEAwatch protocol led to a higher proportion of recurrences being detected by CEA-based blood test and reduced the number detected by patient self-report. This might be because the intensive CEAwatch protocol has more frequent
CEA measurements and less frequent clinical visits than the CAU protocol. That the CEAwatch protocol led to a higher proportion of recurrences being detected by CEA-based blood test is important because those detected by blood test were associated with significantly better survival than those detected by patient self-report, indirectly supporting this intensive protocol. Thus, current evidence confirms the importance of postoperative follow-up, although the optimal protocol remains to be determined.

Colorectal cancer surveillance guidelines differ between countries, and empirical evidence to identify the best surveillance programme remains scanty [10]. The Dutch guideline still advises imaging with ultrasonography of the liver and thoracic X-rays, whereas most other countries recommend CT of the chest and abdomen, although this is not yet proven to be more effective. CEA testing is standard in all countries, but its frequency differs; most western European countries advise CEA tests between 3 and 6 months. Colonoscopic surveillance is recommended once every 5 years after surgery in both Great Britain and the Netherlands. A common recommendation is that follow-up should focus on patients with higher-stage primary tumours, as these are at greater risk of developing recurrence (for example, the European Society for Medical Oncology guideline [16]). However, this does not take into account the outcome of treatment of metastatic disease by primary tumour stage. Specifically, early recurrence after liver resection is associated with high primary tumour stage [21] and, as shown here, survival among patients with a lower primary tumour stage was slightly improved compared with that of patients with stage III tumours. This suggests that survival mainly depends on the biological characteristics of the primary tumour. Following this line of reasoning, it does not matter whether recurrent disease is detected and treated earlier in patients with high-stage primary tumours, because it is the biological tumour behaviour that ultimately lowers the survival probability, with earlier detection having little to no effect on the disease course. Thus, the
higher incidence of recurrences with poorer outcomes at high primary tumour stages must be balanced against the lower incidence of recurrences among low-stage tumours for which treatment is more effective.

The stepped-wedge trial design was statistically challenging. Owing to the sequential roll-out of the intervention, there were legitimate concerns that the majority of patients in the CEAwatch group would already have had a better prognosis because they had survived the CAU period, thereby leading to confounding. However, with the inclusion of patients who had undergone surgery before study enrolment (maximum of 3 years), and the dynamic recruitment during the trial, patient characteristics were considered to be balanced at different phases of the trial. As a result of the cluster design, this study was also prone to an unbalanced distribution of potential confounders. There were age differences between the two groups of patients (Table 5.2). Three centres, with an early switch time, had younger patients. Because of this susceptibility to an unbalanced distribution of potential confounders, these were corrected for in all the analyses.

The effect of the CEAwatch protocol on the transition from being disease-free after surgery (healthy state) to having recurrence detected (illness state) has been described previously by the CEAwatch trial investigators [18]. The present study focused on the transition from detection of recurrence to death. This approach is justifiable under the assumption that the underlying process is Markovian; that is, the probability of death depends only on the patient’s current status.[1, 12] The sensitivity analysis showed that the data were consistent with the assumptions made during the analysis. It is also noteworthy that lead time bias is not of concern in the present analysis, as the full survival time, which was defined as the interval between the date of primary surgery and the event date, was taken into consideration.

The authors propose that quality of life and costs, rather than sim-
ply the primary tumour stage, should be considered when developing an optimal and efficient scheme for the follow-up of colorectal cancer. The motivation for the CEAwatch study was to find a more cost-effective follow-up system, based on the less costly CEA test, which could reduce the number of hospital visits during follow-up by identifying patients who need more than annual imaging. Although this strategy was cost-effective [19], the DSS and OS were comparable.

This study demonstrated no direct survival gain from the intensive CEAwatch follow-up programme. However, this protocol led to a higher proportion of recurrences being detected by a CEA-based blood test and a reduction in the number of recurrences detected by patient self-report. This is an important finding as survival is significantly improved if recurrence is detected while still asymptomatic. The data thus provide indirect support for postoperative follow-up with the new CEAwatch protocol. The authors propose that routine follow-up should now include CEA measurements to identify patients who need more frequent imaging.

ACKNOWLEDGEMENTS

C.J.V. and Z.Z. contributed equally to this work. The authors thank the following study coordinators of the participating hospitals in the CEAwatch trial: E. Manusama, H. van der Mijle, B. Lamme, K. Bosscha, P. Baas, B. van Ooijen, G. A. P. Nieuwenhuijzen, A. Marinelli, E. van der Zaag and D. Wasowicz. They also thank R. Sykes (www.doctored.org.uk) for editorial services. This study was funded by the Netherlands Organization for Health Research and Development (project numbers 171002209 and 171002211). The work of Z.Z. is supported by the Chinese Scholarship Council.

REFERENCES


