Surrogate endpoints in colorectal cancer

G. Browman MD,* A. Shah MDCM,†
J. Ayoub MD,‡ R. Wong MD,§ J. Maroun MD,||
and G. Batist MD# (Colorectal Cancer Clinical Endpoints Working Group)

1. INTRODUCTION

There is ongoing debate about the criteria that are required for a new treatment to be approved for prescription and reimbursement and for clinicians to be convinced of that treatment’s true clinical utility. Although a number of participants are involved in these discussions (some with specific interests), no one denies that the most important interest, that of patients, is served by the rapid availability of safe and effective drugs.

In Canada, the path to new drug availability begins with the formal approval process at Health Canada. Estimates of the average time and cost required to bring a new, effective pharmaceutical to market have stirred discussion about what can be done to improve the efficiency of the drug development process. One proposed strategy is to use alternative endpoints for survival (or, more correctly, mortality as the observable event) in clinical trials that are evaluating new treatments.

In this article, we examine the options for surrogate endpoints in colorectal cancer (CRC) clinical trials, and we examine alternative drug-approval guidelines. Our objective is to provide the consensus view of a representative group of Canadian academic medical oncologists who are involved both in new drug development programs and in the treatment of patients with CRC. The essential questions are whether and how to amend the current guidelines for drug approval to make a wider spectrum of safe and effective treatment options more rapidly available to Canadian patients.

2. ALTERNATIVE ENDPOINTS: TYPES AND USEFULNESS

Clinical trials typically use mortality as the most unambiguous and persuasive primary endpoint in evaluating the effectiveness of an anticancer treatment or intervention. This choice presumably reflects survival prolongation as the outcome that a patient values most when seeking anticancer treatment. The use of mortality also has additional utility: it is easy to measure, and it is resilient against biased assessment. Moreover, its value as a common final outcome reconciles the overall effects of an intervention in terms of benefit and risk.

Table 1 lists the various ways in which a separate clinical trial endpoint could be used with, or as an alternative to, mortality in the evaluation of an anticancer intervention. An alternative endpoint can be used in addition to mortality (to complement mortality as the main outcome of interest) or as a substitute for mortality. When used as a substitute for mortality, the alternative endpoint may be considered the more clinically relevant or important outcome of interest in place of mortality, or, if mortality remains the most important outcome, it may be used as a surrogate for mortality. The use of an endpoint as a surrogate for mortality is implicit acknowledgment that mortality is the most relevant outcome. The surrogate is therefore presumably used to address certain concerns associated with use of mortality as the primary outcome in the evaluation.

Choosing an alternative endpoint as a substitute for mortality in a clinical trial because that endpoint

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Alternative endpoints for mortality in clinical trials of anticancer agents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endpoint type</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>Complementary</td>
<td>The endpoint adds information about the efficacy or effectiveness of the agent and complements information about mortality.</td>
</tr>
<tr>
<td>Substitution</td>
<td>The endpoint is used in place of mortality as the more clinically relevant primary outcome.</td>
</tr>
<tr>
<td>Surrogate</td>
<td>The endpoint is used in place of mortality as its surrogate, with mortality acknowledged as the more clinically relevant outcome.</td>
</tr>
</tbody>
</table>
is more clinically appropriate is more a matter of values than of science, and the choice depends on the question posed. For example, it would be gratifying if either bisphosphonates or erythropoietin alpha reduced mortality in cancer patients, but the primary purpose of these interventions is to address symptoms. Approval of these medicines for marketing should be based on studies of their effectiveness in addressing the main clinical targets of reducing the clinical consequences of bone complications and of anemia, respectively. The effects of these agents on survival are a safety issue. In terms of values, it could be argued that an antitumour agent that prolongs time to disease progression—“progression-free survival” (PFS)—independent of a beneficial effect on mortality is worthwhile, but that argument has nothing to do with using time to disease progression as a surrogate endpoint.

When an alternative endpoint is intended for use as a surrogate for mortality in an intervention trial, the effect on the surrogate of the interventions being tested must obviously reflect differences in subsequent mortality. Because mortality remains the most relevant and unambiguous outcome of interest, the decision to use a surrogate must be motivated by a concern about mortality as the primary endpoint or by some advantage associated with use of the surrogate. If mortality is acknowledged to be the main outcome of interest, concerns about its use usually relate to the efficiency of the trial:

- Participants in the trial must be followed for a very long time before death as the endpoint can be observed.
- The trial requires a large number of patients, resulting in a long accrual period.
- Mortality is relatively infrequent as an event, and thus more patients and longer follow-up are required to detect differences between the treatment arms.
- Co-interventions have or will be used (for example, other agents that have independent efficacy in CRC may confound the interpretation of survival results in the context of the agent being tested).

Table II summarizes the advantages that use of an alternative endpoint as a surrogate for mortality might hold for the efficiency of clinical evaluation of an intervention. If the surrogate endpoint occurs much earlier in the clinical course of the disease being investigated than mortality does, then use of the surrogate would shorten the observation period required to make inferences about the efficacy of the intervention. This savings in observation time for each patient can translate into an overall reduction in the follow-up time needed for the population as a whole. A stark example of this saving occurs in the clinical evaluation of either dietary or pharmacologic inter-

ventions in cancer prevention trials, in which the most appropriate clinical endpoint is new cancer cases. Together, the long delay between exposure to the intervention in healthy people and development of cancer and the relative infrequency of a new cancer in the population make such trials difficult to conduct, highlighting the value of finding surrogate endpoints as time-saving devices.

The use of an alternative endpoint as a surrogate for mortality might also improve the efficiency of clinical evaluation when the surrogate can be assessed based on fewer patients, thus reducing the overall sample size requirement. A smaller sample size can be achieved in two ways:

- Endpoints that are continuous variables (as opposed to discrete variables) permit the comparison of means (as opposed to proportions), and therefore fewer patients are required to detect a relative effect of similar size. An example would be the mean change in CD4 cell count in AIDS intervention trials.
- Endpoints associated with events that occur more frequently than mortality will occur more frequently in a trial, even using a smaller population. A modified version of this approach was used in the ATAC (Arimidex, Tamoxifen Alone or in Combination) trial, which used three endpoints (recurrence, contralateral breast cancer, and death) as eligible “breast cancer events” (the surrogate) to reflect treatment efficacy in the interim analyses. A similar approach has been used in trials to evaluate bisphosphonates, in which the main outcome, “skeletal events,” included multiple possible endpoints.

For evaluators to have confidence about using an earlier-occurring alternative endpoint as a surrogate for mortality, the surrogate must not only be a valid and reliable predictor of mortality, but its predictive ability must also be sufficiently precise to provide information about the magnitude of the clinically relevant difference in mortality between the interventions being compared. Precision about the magnitude is important because clinical, health policy, and business decisions about the use of a new treatment will be affected by how much benefit the new treatment confers and at what clinical cost (toxicity, for example) or financial cost to the patient (clinical decision), to society (policy decision), or to the supplier (business decision). Industry might decide to cease further investment in developing or scaling up production of a product that, although it provides a small mortality benefit, does so with substantial side effects or costs that make the trade-off either clinically or financially unacceptable. Evaluators must also ensure that the motivation behind the use of the surrogate endpoint is not to mask a relatively small difference in mortality that is asso-
TABLE II  Ways in which use of a surrogate endpoint for mortality could improve the efficiency of a clinical trial of an anticancer treatment

<table>
<thead>
<tr>
<th>Characteristic of the surrogate event</th>
<th>Advantage conferred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporally precedes mortality</td>
<td>Allows observation of the surrogate event earlier in the clinical course of the treated disease.</td>
</tr>
<tr>
<td>More frequent than mortality</td>
<td>Allows observation of more events over the same period of follow-up, thus increasing the power of the study.</td>
</tr>
<tr>
<td>Continuous variable</td>
<td>Allows formal comparison of mean effect, which usually requires a smaller sample size to detect similar relative effect sizes, depending on the variance associated with the measure.</td>
</tr>
</tbody>
</table>

associated with apparently larger differences in the event rates of the surrogate.

Validation of an alternative endpoint as a useful surrogate for mortality is a scientific process that usually requires a statistical approach demonstrating a predictable, quantifiable, and usually biologically or epidemiologically credible relationship between the surrogate and mortality. Statistical criteria that have been suggested for making the conclusion that an alternative endpoint is a valid (and credible) surrogate for mortality include these:

- A demonstrated association between the surrogate and the disease
- A demonstrated association between the surrogate and the intervention
- Demonstration that the expected effect of the intervention on the disease is mediated through the surrogate

The next section examines the implications of these considerations in the various clinical contexts of CRC.

3. PROBLEMS WITH SURVIVAL STUDIES FOR ADVANCED-STAGE CRC

The current criteria for new drug approval require clinical trials to provide survival data, which necessarily means a significant sample size and duration of study. Other than these obvious requirements, studies in metastatic CRC that require use of mortality as an endpoint face an additional hurdle. Successes in drug discovery since the mid-1990s have produced many different agents with important clinical benefits for patients. Thus, patients with metastatic CRC entered into clinical trials can receive other active agents, either concurrently or in sequence (that is, in crossover designs) that themselves can affect mortality. Treatment with those agents may mask or obscure the true effect of the agent under investigation.

The difficulties arising from the requirement for survival data when seeking approval for antitumour drugs in CRC are prompting an exploration for alternative endpoints that could be incorporated into current guidelines—whether as surrogates for survival or as important independent outcomes. The consideration of alternative endpoints for CRC is in the interest of patients, health care providers, and the pharmaceutical industry. Alternative endpoints could efficiently expedite the process of drug approval, thus increasing the available options for successful treatment of the disease.

4. ALTERNATIVE ENDPOINTS FOR CLINICAL TRIALS IN METASTATIC CRC

Significant advances in the treatment of many common tumours have been achieved in the past few years. Metastatic CRC usefully exemplifies this recent progress. Although 5-fluorouracil (5-FU) was the only drug available for more than thirty years, the development of new drugs with mechanisms of action other than thymidylate synthase inhibition (for example, oxaliplatin and irinotecan) and of novel combinations [for example, sequential FOLFIRI/FOLFOX (irinotecan–leucovorin–5-FU/oxaliplatin–5-FU/leucovorin)] has provided additional effective second- and third-line therapeutic options. Recent evidence of efficacy with the addition of targeted agents such as bevacizumab or cetuximab to chemotherapy has added further to the potential benefit and complexity of treatment.

These advances are welcome and have led to improved outcomes in the treatment of cancer. However, the evolution of multiple potentially effective agents has also led to new challenges in clearly evaluating the contribution of new agents to the efficacy of individual regimens when overall survival is used as the traditional “gold standard” of efficacy in drug-approval trials. Hence, the need to explore alternative methods—including alternative outcome measures—to help tease out the relative independent contributions of new therapies to overall clinical benefit.

For a surrogate endpoint to be an effective and valid substitute for the main clinical outcome of interest, the effect of the intervention on the surrogate must reliably predict the overall effect on clinical outcome. Thus, a surrogate endpoint must be a correlate of the true clinical outcome, must fully capture the net effect of the intervention on that clinical outcome, and must reflect the magnitude of that net effect. A useful starting point for finding a credible surrogate clinical endpoint for survival is to identify an event that is likely to occur along the same causal pathway from the biologic effect of the intervention to the clinical benefit and that is likely to be associated with mortality. This “event” may be an alternative clinical measure or a biologic marker. In regard
to CRC, an example is change in the level of carcino-embryonic antigen (CEA).

Several investigators have examined the potential roles of tumour response and PFS as valid surrogate endpoints for survival in CRC. These efforts are justified by the notion that, in the absence of tumour response and significant time to progression, a survival benefit is not likely to accrue. The question is then “What is the specific relationship between a given magnitude and rate of tumour response and a given duration to progression that predicts an ultimate survival benefit?”

In addition to predicting changes in the primary endpoint, an ideal surrogate will provide information about the magnitude of change that the intervention will exert on survival based on the magnitude of change that it will exert on the surrogate. Analysis of multiple studies of treatments from among a common class of treatments might help to determine the strength of the correlation between a surrogate endpoint and survival. This hypothetical relationship would need to be tested. Preliminary evidence is currently available for a significant correlation between the common endpoint measures now used to evaluate chemotherapy benefits in clinical trials involving patients with metastatic CRC.

Some potential surrogate endpoints are discussed below with special consideration regarding their application in CRC clinical trials.

4.1 Biologic Markers

Biologic markers of disease are a highly quantifiable measure. They offer potential as surrogate endpoints in CRC trials. The cellular adhesion molecule CEA promotes the aggregation of malignant cells and is also involved in immunity, apoptosis, and metastasis. Serum levels of CEA are elevated in a number of cancers. In metastatic CRC, CEA is elevated in 85% of patients, making it the marker of choice for monitoring CRC. Even in the absence of objective response, CEA levels correlate with survival; moreover, the relationship between a CEA response and a true objective response is significant for tumours that secrete CEA.

Problems remain in regard to the significant number of tumours that do not express or that fail to secrete this biomarker, and to well-documented examples of discordance (false positives and false negatives) between radiographically confirmed response and change in CEA level. The true value of CEA as a potential surrogate for survival in metastatic CRC will require a population–based study in which the overall performance of the test is evaluated for its sensitivity and specificity and for predictive values of CEA in relation to mortality. A study of this sort has not been performed in the past, likely in part because of the limited efficacy of systemic treatments in CRC, so that the value of blood levels of CEA or of changes in blood levels as surrogate endpoints remains undetermined.

Other biologic markers that better predict mortality may be developed in future. Recent developments in proteomic analysis in blood have suggested this possibility. Because response to any individual treatment is usually observed in only 25% – 30% of patients, the discovery of biomarkers that identify tumours or patients likely to respond or benefit is an area of great activity that holds the promise of greatly accelerating clinical trials by focusing on subpopulations enriched for a particular biologic phenotype.

4.2 Response Rate

Traditionally, change in the size of a tumour has been used to define that tumour’s response to therapy. Response rate (RR) has been the primary surrogate endpoint supporting accelerated approval for cancer drugs. Studies are conducted in patients with tumours that have already been proved refractory to available therapy or that have no effective therapy. Thus, a single-arm study demonstrating an objective RR of acceptable magnitude and duration could be seen as preliminary evidence of drug activity that could confer benefit. In comparative controlled studies, RR data could serve as a suitable surrogate for survival if the magnitude of the effect of the intervention on RR can be correlated with a reasonably narrow range of measures of overall survival (OS) and if patients are suitably stratified.

For example, patients with long-lasting stable disease may progress much later than those who attain a complete response of short duration. Buyse et al. analyzed pooled individual patient data from 3791 patients entered into twenty-five randomized clinical trials of first-line novel treatments compared with 5-FU–based chemotherapy in metastatic CRC. The pooled analysis detected an overall average improvement in OS for an 11.4% increment in RR.

That correlation is described by the regression equation

\[ \text{OS (months)} = 0.088 \times \text{RR} \times 10.45. \]

That equation translates into a 1-month increase in OS for an 11.4% increment in RR.
4.3 Progression-free Survival

Although OS is considered the “gold standard” in the approval and clinical use of new therapeutic agents, the other outcome that has been studied as a potential surrogate is PFS, usually measured from the time of randomization.

The argument has been made that PFS is an adequate surrogate endpoint for survival in the treatment of CRC in both the adjuvant setting and the context of metastatic disease. The definition of “progression” in PFS is important for determining the accuracy of the endpoint. Several aspects of a study design that uses PFS as an endpoint can influence the interpretation of the trial results. For example, the evaluation method used and the schedule of follow-up can lead to ascertainment bias. Ideally, progression should be determined based on an objective measure such as, in metastatic CRC, radiographic confirmation of changes in tumour size and, in the adjuvant setting, the first appearance of recurrence or metastasis. Progression based on worsening symptoms or a change in a biologic marker may also be used, but these are subject to measurement bias, especially in trials without placebo controls.

A principal advantage of PFS over OS as a clinical endpoint for evaluating the efficacy of a drug is that it potentially allows for determination of the efficacy of the initial therapy even in a crossover design, when patients receive further treatment after tumour progression. Crossover designs preclude the use of survival as a meaningful endpoint because the cross-over masks differences in efficacy between the control and the experimental treatment arms. With the availability of several active drugs for CRC, masking is a clinically relevant issue.

In the Louvet study mentioned earlier, a significant correlation was also found between PFS and OS ($r = 0.481, p < 0.0001$). The relevant regression equation is

$$\text{OS (months)} = 0.68 \times \text{PFS (months)} + 8.74.$$  

Thus, a 1-month increase in PFS corresponds to a 0.68-month increase in OS.

4.4 Time to Tumour Progression

Time to tumour progression (TTP) from the date of randomization or the date of treatment until the time of documented progression or death is often used to define PFS. The TTP is primarily based on the investigator’s assessment of the tumour based on standard objective response criteria such as those put forward by the World Health Organization or those of the Response Evaluation Criteria in Solid Tumors.

When TTP is used in outcome evaluation, important issues have to be considered if bias is to be avoided and the validity of TTP as a surrogate for survival ensured:

- The minimum interval between tumour assessments should be less than the expected size of the treatment effect on TTP.
- The frequency of tumour assessment should be the same across all study arms even when treatment cycles are of different lengths.

Miller et al. presented a retrospective analysis of pooled data from two randomized controlled trials that tested irinotecan against 5-FU-based chemotherapy as first-line treatment in metastatic disease. They reported that both trials individually detected significant improvements in both TTP and OS. To exploit this concordance in outcomes, the investigators used regression analysis to derive a statistical formula describing a linear relationship between TTP and OS that was independent of treatment or baseline prognostic factors as covariates. In a Cox regression analysis, a TTP $\geq$ 6 months was a better predictor of OS than was any other prognostic factor. The abstract does not provide sufficient information about length of survival to know how much time would be gained by making an inference about effectiveness at 6 months based on the TTP rather than waiting until mortality can be assessed.

To study the relationship between OS and intermediate endpoints that could potentially substitute for OS, Di Leo et al. analyzed pooled individual patient data from four separate randomized trials that compared first-line 5-FU-based therapy with 5-FU-based therapy plus an investigational agent (either irinotecan or oxaliplatin). In two of the trials (testing irinotecan), response was the primary endpoint. In the other two trials (testing oxaliplatin), tumour progression was the primary endpoint. No trial was sufficiently powered to detect plausible OS differences. In general, under traditional “best research practice,” using analyses to make up for such deficiencies in trial design is frowned upon.

In the two trials testing irinotecan, OS and the intermediate outcomes RR and TTP were all improved; but, in the two trials testing oxaliplatin, OS was not improved despite improvements of similar magnitude in the intermediate endpoints. The authors were unable to confirm any of their hypotheses explaining the lack of OS benefit for the oxaliplatin trials. (Of course, this is why randomized trials are conducted and designed to achieve internal validity.)

Assuming that the use of a surrogate implicitly acknowledges mortality as the most appropriate endpoint, the foregoing example argues against rather than for using surrogates for mortality. On the other hand, it does pose the question of whether an alternative intermediate endpoint ought to be considered the more clinically relevant one, and therefore used.
in place of mortality to influence policy decisions about the commercialization of a product.

To evaluate whether PFS correlates with RR and OS, Louvet et al. analyzed pooled data from all phase III studies published between 1990 through 2000 regarding first-line treatment for patients with metastatic CRC. A total of 13,498 patients from twenty-nine studies with sixty-five treatment arms (each with more than 100 patients) were included in the analysis. Louvet and colleagues detected a significant correlation between PFS and RR \((r = 0.655, p < 0.0001)\). That correlation is described by the regression equation \(y = bx + a:\)

\[
PFS \text{ (months)} = 0.1 \times RR \text{ (%) } + 3.2.
\]

Thus, for each 10% increment in RR, PFS increased by 1 month. As noted earlier, a correlation was also evident between RR and OS \((r = 0.408)\) and between PFS and OS \((r = 0.481)\). The strongest correlation occurred between RR and PFS, indicating that a change in RR is associated with a greater change in PFS than in OS. The smaller correlations between RR and OS and between PFS and OS can be accounted for by the fact that RR and PFS reflect the effects of first-line treatment alone, but that OS is more likely to reflect the effect of first-line treatment and all subsequent therapeutic interventions. These results are consonant with the view that PFS may be a more sensitive measure of treatment efficacy than OS is, especially in patients who receive sequential treatments.

4.5 Working with Candidate Surrogate Outcome Measures

Based on the evidence thus far, either TTP or PFS seems to be the most promising candidate surrogate outcome measure with which to evaluate the potential benefits of a new agent in CRC. The data from Louvet et al., together with the suggestive supporting evidence from other attempts to validate one or the other endpoint as a surrogate, seem to provide sufficient grounds for prospective validation at least of PFS or TTP as a useful surrogate for survival in future studies of new agents in CRC.

Given the promise of many emerging agents and the clinical consequences of CRC to patients, it would seem reasonable to support an agenda of prospective validation of surrogate endpoints for trials of new agents in CRC to finally answer this question. Regulatory agencies may wish to use the opportunity of a prospective approach of this kind to make approval decisions based on the effects of new agents on a surrogate endpoint, pending further clarification as the studies mature.

Such an approach would represent a productive collaboration between investigators, regulators, and industry to address an important problem in evaluating new agents in the context of CRC, where use of survival as the main outcome may be imposing a burden of proof that is unreasonable and that may deprive stakeholders of potentially beneficial treatments.

If such an approach were to be adopted, it would be important to ensure that standards for conducting trials using TTP or PFS be developed and published, that investigators adhere to the standards, that industry support and facilitate the use of the standards, that all data emanating from such studies be made available for analysis, and that all parties be prepared to either accept or reject the use of such a surrogate in future depending on the results of this unique collaborative effort.

5. ALTERNATIVE ENDPOINTS FOR CLINICAL TRIALS IN ADJUVANT CRC

Traditionally, OS has been considered the most important outcome in evaluating adjuvant therapy in CRC. Using this endpoint, several adjuvant treatment strategies have been established to be effective in stage III patients who are at high risk of recurrence. Use of survival as the primary endpoint has required prolonged observation of study participants in the evaluation of potentially curative regimens. In contrast to the relatively shorter observation period required to observe mortality as an event in metastatic disease, an observation period of more than 8 years may be required to accumulate evidence for an effective treatment in the adjuvant setting.

Disease-free survival (DFS) could potentially serve as a useful and valid surrogate endpoint for survival, providing speedier evaluation of new adjuvant therapies. Disease-free survival refers to the period during which a patient has no clinical evidence of cancer; it may therefore also be considered a clinically meaningful outcome in its own right.

In a meta-analysis of individual patient data from seventeen randomized controlled trials involving 17,367 patients, Sargent et al. demonstrated a strong association between DFS and OS in patients receiving 5-FU–based adjuvant chemotherapy. The event rates for 3-year DFS and 5-year OS were nearly identical, meaning that using 3-year DFS as a surrogate endpoint would not necessarily reduce the sample size required to achieve statistical significance, but it could produce efficiencies by reducing the observation period. However, preliminary findings indicate that marginally significant 3-year DFS may not translate into significant 5-year OS benefits. That observation challenges the utility of DFS as a valid surrogate for OS, but it also raises the question of whether DFS on its own ought to be considered a clinically important measure of treatment benefit. Of all CRC recurrences, 75% occur within 3 years after diagnosis, and most of those recurrences are incurable. Therefore, designing adjuvant chemotherapy studies using 3-year DFS as the endpoint is logical. Further
evaluation is required to determine the utility of 3-year DFS as a surrogate for survival.

Based on the available information, the Oncologic Drugs Advisory Committee of the U.S. Food and Drug Administration (FDA) recommended in May 2004 that, as compared with results with standard therapy, an increase in (at least) 3-year DFS represents clinical benefit and is adequate grounds for regular drug approval. The 3-year DFS is now the primary endpoint of the next generation of CRC adjuvant studies.

6. DISCUSSION

To be convinced of the clinical benefit of a new drug for the treatment of CRC, regulatory agencies and physicians currently require survival data before they will offer that drug to patients. This requirement is highly rigorous, but it demands long study periods that add significantly to the cost of evaluation and to the time needed before a potentially beneficial agent can become available to patients. The use of surrogate endpoints in CRC clinical trials should seriously be considered as an attractive solution to this problem.

Although we generally agree with the need to consider alternative clinical endpoints, caution must be exercised in choosing a suitable surrogate endpoint in a clinical trial. Using a surrogate endpoint for survival has the risk of false-positive or false-negative inferences about the effectiveness of the intervention. This risk might be most acceptable where the magnitude of the savings in time or cost of the evaluation are very large—such as in cancer-prevention trials. In the adjuvant setting, the time delay between the first event (recurrence) and death may be long enough to warrant use of a surrogate endpoint (for example, disease recurrence). In the context of the treatment of metastatic disease, some of the proposed substitute endpoints may occur in reasonably close temporal proximity to death (for example, disease progression or recurrence versus death in colon cancer), thereby minimizing any benefit from using the surrogate. However, with the advent of several new drugs that have clinical efficacy for CRC, this last issue has become somewhat more complex. Patients with CRC are now living significantly longer and are often candidates for multiple sequential therapies. This clinical practice can blur the independent contribution of individual treatments to overall patient survival. It would therefore be useful to identify criteria for the introduction of surrogate endpoints for mortality in each of these contexts.

Concerns also arise about how bias might be introduced into trials that use surrogates for mortality. Bias can arise from two sources. First, in non-blinded trials, patients randomized to an experimental or control treatment may be managed in subtly different ways because of inherent, unconscious biases associated with the knowledge of the patient’s treatment assignment. The result could be bias associated with differential application of co-interventions, possibly including frequency of follow-up or intensity of diagnostic intervention upon suspicion of a clinical problem. That type of bias can occur whether mortality or a surrogate is used. However, the timing of death as an observed event is difficult to manipulate; surrogate events, such as disease recurrence or tumour response, are affected by variations in the frequency and intensity of observation. Also, mortality is not affected by observer variation, but considerable within- and between-observer variation has been documented in the evaluation of tumour response.

A move to surrogate endpoints must therefore be accompanied by greater rigour and an acknowledgment that the introduction of surrogates enhances the risk of increased random and systematic error in the assessment of outcomes.

A second potential source of bias in using a surrogate for mortality in an intervention trial is the problem of multiple comparisons that might actually neutralize any potential benefits for using the surrogate. It would be naïve to think that a clinical trial that is failing to detect a difference in the surrogate endpoint in favour of an experimental intervention would not be further followed in the hope that survival differences might emerge, especially if interim trends are observed. In such a situation, the trialists will be tempted to claim a benefit for the intervention, and more importantly, consumers will be convinced of the potential benefits of the treatment. However, to be able to conclude in such a circumstance that a treatment provides benefit, the trial would have had to be designed to adjust the size of the type I error (alpha) to protect against such false inferences. In ordinary language, by using a surrogate endpoint, but by also being prepared to claim success of the intervention based on mortality, the investigator and sponsor are increasing the likelihood of observing a desired result by chance alone. This problem is related to that of multiple comparisons. The required adjustment in alpha for calculation of the sample size might, in many circumstances, neutralize the efficiency gains envisioned from use of the surrogate endpoint. This problem would likely cause regulators to insist on methodologic strategies to protect against such biases.

Investigators observing important clinical differences in a surrogate endpoint between an experimental and a control intervention in a trial will want to conclude that such differences reflect (or predict) differences in mortality, assuming that the alternative endpoint has been shown to be a valid surrogate. However, in the case of a new investigational agent, there is never certainty that the intervention has no unexpected effect that could actually be harmful to patients in a way that reduces mortality. Thus, for
safety reasons, mortality data may often be required anyway. Three well-known examples of this phenomenon are

- the recent revelations concerning the increased incidence of cardiovascular deaths associated with use of some COX2 inhibitors,
- the failure of clofibrate to improve overall survival because of an increase in competing risks of death despite clofibrate’s beneficial effects on blood cholesterol levels, and
- the unexpected increase in mortality after myocardial infarction associated with use of a certain class of anti-arrhythmia drugs despite the observed benefits of those drugs on cardiac arrhythmias.

Most of these cases have come to light after long-term use of the agent in question. This concern is of particular note in the case of adjuvant therapy for CRC, but also in the context of metastatic CRC, in which median survival is now significantly extended to about 20 months. In any event, post-marketing surveillance is being greatly enhanced as a result of such examples and would no doubt be extended to anticancer agents approved on the basis of surrogate endpoints.

Although mortality might still need to be documented for the foregoing reason, the choice of a surrogate endpoint as the primary endpoint in a trial could reduce the overall sample size required, because sample sizes are based on the primary outcome of interest. However, investigators would have to be fairly sure that such sample sizes are adequate to credibly explain important trends in the mortality data that could flag potential concerns.

Many cancer care products have side effects in addition to their intended effects. Many of the frequent side effects are predictable and can be deduced in advance based on the known biologic mechanisms of action of the drug; but many examples exist of unanticipated side effects and longer-term consequences of anticancer and other therapies. Even if an intervention improves survival, patients and practitioners will always want to know whether the magnitude of the benefit for a particular patient outweighs the risk of adverse events. Shortening the observation period, and therefore the number of patients required in a clinical trial (assuming that fewer would be needed to observe the surrogate outcome), could result in missed opportunities for evaluating some of the important unintended effects. As a result, concerns about due diligence in the drug development model could arise, risking litigation in cases where surrogate endpoints are used.

One solution might be to insist that the number of patients entered into a trial be calculated to attain the equivalent in person-years of follow-up that would have occurred for a reasonably powered study had mortality been the outcome of interest. Through such a mechanism, the overall time of observation could still be reduced, but through insistence on a higher-powered study that would create sufficient events over the shorter observation period to detect the surrogate outcome, sufficient observations will be available to detect less frequent adverse events (although still not all of them). Such an approach would at least signal concern about due diligence in drug evaluation. However, the strategy would work only when event rates are either constant over time or more frequent earlier in the follow-up period.

7. RECOMMENDATIONS

7.1 Clinical Recommendations

7.1.1 Metastatic CRC Therapy

With respect to metastatic CRC, we favour the use of TTP as a valid surrogate endpoint for clinical trials. Among the currently available options, TTP offers the most objective, reliable, and practical alternative to survival and symptom-control endpoints. It represents the most common cause of treatment failure, incorporates time into the value, and offers a direct assessment of disease burden that logically correlates with symptom progression and survival.

Clinical trials based on surrogate endpoints are inherently less powered to categorically confirm the clinical benefit of an intervention. We recommend that trials based on TTP be evaluated to grant accelerated approval of drugs in CRC. Regular approval of the drug may be subject to the completion of phase IV trials on more conventional measures of clinical benefit. The completion of post-marketing trials can be promoted by incorporating a requirement for initiation of the trial before the drug is approved. Alternatively, conditional approval may be granted based on interim analysis of TTP in a randomized trial, with clinical benefit to be confirmed upon the trial’s completion. If this strategy were to be followed, then clear decision rules would need to be established to address the problem of entitlement, because of the difficulty involved in cancer agencies and governments withdrawing already-established treatments based on post-marketing studies. However, recent experiences with serious adverse events identified long after approval of COX2 inhibitors demonstrates that withdrawal can be accomplished. Such an approach will require a more formal appreciation of the nature of approval based on surrogate endpoints being “tentative” and subject to significant change depending on follow-up data.

7.1.2 Adjuvant CRC Therapy

In the adjuvant setting, we would favour the adoption of DFS with at least 3 years of follow-up as the preferred surrogate endpoint for CRC trials. This recommendation is in keeping with the current approach...
of the FDA and of current adjuvant studies. Adoption of this endpoint will allow trials to be completed and agents to be incorporated into practice sooner.

7.2 Methodology Recommendations

We recommend that a policy of adopting alternative endpoints as surrogates for survival in trials of CRC be accompanied by close examination of the associated methodologic issues as raised in this report and by design and development of formal methodology standards for conducting and reporting such studies.

7.3 Policy Recommendations

We recommend that the adoption of alternative outcomes as surrogates for survival for the purpose of increasing the efficiency of evaluation be accompanied by close examination of the pricing policies for new agents, so that cost savings resulting from such studies can be fairly distributed between payers and shareholders.

We further recommend that adoption of endpoints as surrogates for survival be accompanied by the development of clear rules governing decisions about the continued availability of agents based on the results of post-marketing studies.

8. CONCLUSION

Considerable progress has been made in the management of CRC over the past few years. Several new active agents have been identified, and patient survival has been significantly extended. This gratifying success has brought new challenges, however. How to evaluate the clinical benefits provided by new treatments in an era in which an increasing proportion of the patient population is receiving multiple sequential therapies calls for a rethinking of suitable clinical endpoints. This rethinking is necessary to ensure that practitioners fully understand the contribution to overall survival of each component of treatment. We encourage open discussion of the relationship of the costs for the classical clinical trial approval paths and the relatively longer times to approval by Canadian regulatory agencies on the one hand, and of the benefits to patients from quicker access and to society with regard to drug costs. Although this paper discusses the issues in CRC, we believe that these same issues have broader applicability in other types of cancer, and we propose that amendments to the evaluation of new cancer treatments provide an opportunity to benefit all stakeholders.

9. ACKNOWLEDGMENT

This program is supported through an educational grant from Bristol–Myers Squibb.

10. REFERENCES

23. Peterson L. FDA Oncologic Drugs Advisory Committee (ODAC) meeting on clinical trial endpoints in colorectal cancer (CRC); Gaithersburg, MD; May 4, 2004 [online newsletter]. Trends in Medicine. May 2004. [Available at: www.trends-in-medicine.com; cited November 24, 2005]

Correspondence to: Gerald Batist, Department of Oncology, McGill University and Montreal Centre for Experimental Therapeutics in Cancer, Sir Mortimer B. Davis–Jewish General Hospital, 3755 Cote St-Catherine Road, Montreal, Quebec H3T 1E2. E-mail: Gerald.Batist@mcgill.ca

APPENDIX

Current Drug Approval Regulations for Colorectal Cancer

Health Canada’s Therapeutic Products Directorate refers to guidance from the International Conference on Harmonization (ICH) for clinical trials seeking drug approval. The ICH guidelines require that a primary endpoint provide a valid and reliable measure of some clinically relevant and important treatment benefit in the patient population based on accepted norms and standards. The importance of experience with endpoints in published literature is emphasized.

The ICH recognizes the need for consideration of surrogate endpoints when direct assessment of the clinical benefit to the subject through observation of actual clinical efficacy is not practical. Two key concerns with the use of surrogate endpoints in clinical trials are expressed. First, the surrogate might not be a true predictor of the clinical outcome of interest. Examples are provided of situations in which positive results using surrogates were not borne out on subsequent examination of survival, and where a positive survival benefit was not observed in preliminary and surrogate endpoints. Therefore, risks of both false positives and false negatives exist. Second, the surrogate endpoint may not yield a quantitative measure of clinical benefit that can be weighed directly against adverse effects. Although statistical criteria for validating surrogate endpoints have been proposed, experience with their use is relatively limited.

The ICH outlines three main factors on which the strength of a surrogate endpoint will depend:

- Biological plausibility of the relationship between the surrogate and clinical outcome
- Demonstration in epidemiology studies of the prognostic value of the surrogate for the clinical outcome
- Evidence from clinical trials that effects of treatment on the surrogate correspond to effects on the clinical outcome

In the United States, the two routes to request drug approval from the Food and Drug Administration (FDA) are regular approval and accelerated approval. Regu-
lar (“full”) approval requires demonstration of either clinical benefit (for example, prolongation of survival or improvement in tumour-related symptoms) or a beneficial effect on an already-established surrogate endpoint. Acceptance of a surrogate for regular approval involves ample clinical experience, with widespread acceptance by both physicians and patients.

Accelerated approval can be given to drugs that are intended to treat serious or life-threatening diseases (of which colorectal cancer could be considered an example) if the new drug appears to provide benefit over existing therapy. Accelerated approval can be based on substantial evidence from well-controlled trials using a surrogate endpoint that is considered reasonably likely to predict clinical benefit. In other situations, such surrogates have included tumour response rate, time to tumour progression, and quality of life. Following accelerated approval, the drug must undergo post-marketing studies within a reasonable time to confirm clinical benefit and remain on the market.

The FDA usually requires more than one trial before approving a drug, but exceptions are made generally only in cases in which a single multicenter study of excellent design provided highly reliable and statistically strong evidence of an important clinical benefit, such as an effect on survival, and a confirmatory study would have been difficult to conduct on ethical grounds.” A single study is also adequate for the approval of additional marketing indications for previously marketed cancer drugs and for additional indications in different stages of the same cancer.

In general, drug approval requires that the drug has a “clinically meaningful effect,” which is often interpreted as an increase in survival as compared with existing treatment. Relevant endpoints should also be clinical events that are important to the patient—that is, an event such as tumour response or an event that the patient is aware of and wants to avoid, such as recurrence. For most patients, some evidence for a possible survival benefit is important, although shrinkage of the tumour (tumour response) and duration of relapse-free progression are possibly very important endpoints for patients even in the absence of convincing survival data.