The Swedish Council on Technology Assessment in Health Care (SBU) Report on Cancer Chemotherapy—Project Objectives, the Working Process, Key Definitions and General Aspects on Cancer Trial Methodology and Interpretation

Peter Nygren and Bengt Glimelius for the SBU-group[1]

From the Department of Oncology, Radiology and Clinical Immunology, University Hospital, Uppsala (P. Nygren, B. Glimelius), Sweden

Correspondence to: Peter Nygren, Department of Oncology, Radiology and Clinical Immunology, Section of Oncology, University Hospital, SE-751 85, Uppsala, Sweden. Tel: +46 18 611 49 41. Fax: +46 18 611 55 28. E-mail: peter.nygren@medsci.uu.se

Acta Oncologica Vol. 40, No. 2/3, pp. 155–165, 2001

The major objective of the Swedish Council on Technology Assessment in Health Care (SBU in its Swedish abbreviation) Chemotherapy Project was to review and assess the evidence, as presented in the scientific literature, for the use of chemotherapy in some major tumour types. The purpose of this assessment was to establish the current evidence-based role of this treatment modality in various settings in these tumour types. The basis for the assessment was the perspective of the conditions in Swedish health care and a prospective survey within the project detailed the use of chemotherapy in two health-care regions in Sweden. This formed the basis for an assessment of the use of chemotherapy in relation to scientific evidence. The project focused on standard cytotoxic chemotherapy, whether delivered at conventional doses or at higher doses with various supportive actions, and did not address the role of other emerging approaches that could broadly be included under the heading chemotherapy, e.g. drugs with anti-angiogenesis and anti-metastatic properties or chemopreventive drugs. In this introduction, the project objectives, working process, key definitions and general aspects on cancer trial methodology and interpretation are presented and discussed.

OBJECTIVES

The literature review aimed at addressing two main questions:

(i) How strong is the scientific evidence that chemotherapy is beneficial in various settings in the treatment of the particular tumour type?

(ii) If there is scientific support for benefit, how large are the effects?

In a second step, the project also aimed at evaluating whether the current use of chemotherapy in Swedish health care is in line with that concluded to be scientifically justified from the literature overview. This was done by a prospective survey within the Stockholm and Uppsala/Örebro regions of the use of chemotherapy during a four-week period in the autumn of 1997, the results of which were then compared with the conclusions from the overview.

The tumour types included in the overview were mainly those where three investigations had shown an increased use of chemotherapy during the latest decade. The first investigation was done prior to a consensus conference organised by Medicinska forskningsrådet (MFR) and Sjukvårdens planerings- och rationaliseringsinstitut (SPRI) in Stockholm in 1990 (1), the second a follow-up investiga-

[1] Other members of the SBU-group were: Jonas Bergh, Radiumhemmet, Stockholm; Lars Brandt, Dept of Oncology, University Hospital, Lund; Bengt Brorsson, SBU, Stockholm; Barbro Gunnars, Dept of Oncology, University Hospital, Lund; Larsolof Hafström, Dept of Surgery, University Hospital, Umeå; Ulf Haglund, Dept of Surgery, University Hospital, Uppsala; Thomas Högberg, Dept of Gynaecological Oncology, University Hospital, Linköping; Karl-Gunnar Janunger, Dept of Surgery, University Hospital, Umeå; Per-Ebbe Jönsson, Dept of Surgery, Helsingborgs lasarett, Helsingborg; Göran Karlsson, Handelshögskolan, Stockholm; Eva Kimby, Dept of Haematology, University Hospital, Huddinge; Gunilla Lamnevik, SBU, Stockholm; Sten Nilsson, Radiumhemmet, Stockholm; Johan Permert, Dept of Surgery, University Hospital, Huddinge; Peter Ragnhammar, Radiumhemmet, Stockholm; Sverre Sörenson, Dept of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway.

 ISSN 0284-186X

tion performed in 1993 in order to study a possible influence on routines by the consensus statement (2) and the third performed during the planning phase of the present report in 1996. Questionnaires were then sent to all departments of oncology, surgery, including urology, gynaecology and internal medicine, including haematology and pulmonary medicine. These showed that the use of chemotherapy had increased as had also the positive opinion of benefit from chemotherapy in non-small cell lung cancer, gastric cancer, pancreatic cancer, colorectal cancer and urinary bladder cancer. Secondly, the two groups including the highest numbers of patients treated with chemotherapy in Sweden, i.e. breast cancer and haematological malignancies, were included. Among the haematological malignancies, only the most prevalent subgroups were chosen for the literature review, i.e. acute myeloid leukaemia, chronic lymphocytic leukaemia, Hodgkin's disease, aggressive non-Hodgkin's lymphoma (NHL), chiefly large B-cell lymphomas, and indolent NHL of the follicular subtype. Taken together, these subgroups constitute about 75% of the haematological malignancies. Finally, ovarian cancer was included since chemotheray is extensively used and, at the time of planning of the overview, one group of very expensive cytotoxic drugs, the taxanes, had preliminarily shown favourable results, and the use of the drugs had started to increase in Sweden. In each tumour type, seldomly occurring clinical presentations and rare morphologic subtypes were excluded. These exclusions are detailed in each diagnosis chapter. Treatment in which chemotherapy is concomitantly mixed with radiotherapy to achieve potentiation of the radiation effect were included only exceptionally.

INTRODUCTION TO CHEMOTHERAPY AND SOME COMMENTS ON CHAPTER CONTENTS

Chemotherapy is mostly administered systemically via the oral or, more commonly, the intravenous route. There are several different settings in which chemotherapy might be used. In a few types of advanced cancer, chemotherapy might eradicate essentially all tumour cells and cure the patient. Chemotherapy is also increasingly used as an adjunct to surgery or radiotherapy, i.e. adjuvant treatment, with the aim to eradicate microscopic disease to achieve cure in early cancer. Administration of chemotherapy prior to surgery or radiotherapy, i.e. preoperative and neo-adjuvant treatment, with the aim to decrease the tumour volume and allow for more effective local treatment and improvement of the overall tumour control, respectively, is also increasing. Finally, palliative chemotherapy is used in advanced cancer not amenable for cure, with the aim to achieve tumour remission followed by symptom control and prolongation of survival.

In addition, chemotherapy can be given locally into body cavities harbouring tumour growth or loco-regionally, e.g. via the hepatic artery into the liver, with the aim to achieve high drug concentrations and a better effect. The types of chemotherapy to be covered in each diagnosis-specific chapter were those considered to be in common use in Swedish health care, to be in focus of recent and current investigation, e.g. high dose chemotherapy, or less commonly used but under prior extensive development.

Since chemotherapy in many tumour types has not a curative potential but might only produce symptom relief and life-prolongation, but essentially always at the expense of more or less severe adverse effects, the issue of quality of life (QoL) during and after chemotherapy has come much into focus during the last decade. It is quite clear from the diagnosis-specific chapters that the extent to which QoL has been addressed differs considerably between tumour types. However, many of the chapters address this important issue and in general, a number of conclusions on QoL are drawn. Since the meausurement of QoL is difficult from a number of methodological aspects, the report includes a separate chapter critically covering this area (3).

Adverse events from chemotherapy was not covered by a systematic review. However, the Swedish version of this report includes a fairly comprehensive overview of this aspect of chemotherapy.

The cost of care is another important issue which is gaining increasing attention in the current situation of financial deficits. Measurement of cost of care, especially when put in relation to what might actually be gained, is also methodologically difficult and is addressed separately in the chapter on cost of chemotherapy (4).

OUTLINES OF THE WORKING PROCESS WITHIN THE PROJECT

A national expert, or in a few cases experts, was selected for each tumour type for searching, assessing, summarising and classifying the literature within the field as well as for writing an initial version of the diagnosis specific chapter. The common outline to be used included an introduction setting the current standard therapy for the tumour type followed by the overview of the literature divided in subsections depending on various treatment settings. At the end, each sub-section was to be summarised, addressing the conclusions that could be supported by the literature and those that could not. In most chapters a table classified and evaluated the literature for the sub-section with respect to type of study, scientific quality and number of patients included (see below).

Treatment results were to be detailed with respect to point estimates of the treatment effects as well as levels of statistical significance for comparisons between treatments. However, as discussed further below, various types of estimates of treatment effects in line with the reporting in the original scientific literature had to be accepted despite

making comparisons between studies/diagnoses more difficult. Furthermore, although the aim was to report the treatment results in the different diagnoses at identical levels of detailing, this was not fully achieved, partly due to variability in the results reported in the literature and partly due to differences between the styles of writing.

When a preliminary draft of a chapter was finished, it was reviewed by a second expert within the SBU project group. Furthermore, each chapter was sent to each member of the project group for additional comments and suggestions for amendments and were discussed at several meetings within the group. The chapters were then mailed to, in most cases two, international experts within each field and some of the chapters were also reviewed by members of the SBU Expert Committee. The additional non-diagnosis specific chapters with less specific and more general scopes were not submitted to international reviewers but were assessed and commented upon within the group and by experts within the SBU.

From this commentary process, many valuable viewpoints were obtained together with considerable complementary information, much of which was incorporated in the document. Some comments, however, reflected preferences and conditions beyond the scope of the literature review and were, therefore, not included.

After considering all viewpoints, each chapter was revised, linguistically reviewed, discussed within and finally approved by the project group. Finally, the SBU Expert Committee, together with 4 external experts, evaluated the complete report. Appropriate viewpoints were finally incorporated in the present final version of the report.

In essence, the procedure used for reaching the final wordings of the manuscripts was aimed at being 'evidence-based' and 'systematic' (5, 6).

PRINCIPLES FOR LITERATURE SEARCH

Literature search, classification, interpretation and integration of evidence were performed along the lines described by SBU (7).

The medical databases Medline and Cancerlit were used to search the literature for reports on clinical trials to form the basis for the literature overview for each tumour type. The general search terms used were the diagnosis, '(randomised) clinical trial' and 'chemotherapy' and the period covered varied, although it covered at least the latest 15 years and for controlled trials at least 20 years. In some tumour types, it was important to cover the entire chemotherapy period starting in the 1950s in order to be able to study the documented evidence for, e.g. survival prolongation. Expertise at the SBU assisted in the primary literature search for all diagnosis-specific chapters.

Additional trial reports and study data were obtained by reviewing the reference lists of relevant articles and adequate reports from the 'gray literature' (5, 7), in most cases symposium proceedings, were identified by the experts primarily involved in chapter writing and by the international experts. The PDQ database, issued by the US National Cancer Institute, as well as current issues of comprehensive oncology textbooks, were also used to check for additional study reports. Thus, in practice, a combination of search strategies was used, a principle used in many systematic reviews (5). Each section of the final report has been updated to include *all* publications through 1998. Key publications until September 2000 of randomised studies already known to be in progress during 1998 were also included.

A general impression among the writing experts was that the computer-based literature searches using the established data bases were not as comprehensive as expected with respect to gain of pivotal information. Thus, personal knowledge of all experts involved, use of reference lists in published articles as well as the 'gray' literature provided additional important information. With respect to the gray literature, abstracts from symposia proceedings were the most important sources and had to be accepted provided that the study data were from established research groups, seemed to be methodologically sound, and presumed to be available as full reports within a limited time frame. Despite the obvious difficulties to assess the limited information provided in abstracts, the alternative to leave all abstracts out was not considered acceptable since many important large trials are often reported at major conferences well in advance of the full publication. As far as time-limits allowed, abstracts were replaced by full reports in the final versions of the chapters.

PRINCIPLES FOR CLASSIFICATION AND ASSESSMENT OF THE LITERATURE

The system based on SBU recommendations (7) was used to evaluate and classify the scientific reports. The system is comprised of two components:

(i) Classification of the report by type of study.
(ii) Assessment of scientific quality to indicate the weight of evidence provided by the report.

Classification was done into the following types of studies:

- *Meta-analyses (M).* A statistical synthesis of several randomised trials. The analysis can be based on original data from the trials or solely on information reported in the literature.
- *Phase III trials (C).* Prospective trials that randomise among treatment alternatives. May be broadly designed and involve the participation of multiple centres or may be conducted at a single centre. The quality of these trials heavily depends on sufficient dimensioning and the internal and external validity of the trial results.

- *Phase I/II trials (P).* Prospective, mostly non-randomised, trials to study efficacy and safety of a new treatment. May include very few patients.
- *Retrospective studies (R).* Studies that analyse retrospective patient data, mostly from one centre or based on international databases collecting information from different centres, that have been collected over a longer period. Treatments have often changed over time. Large, closely-monitored patient studies collecting data according to a predefined protocol, with long follow-up times and presented with adequate statistical analyses can be informative.
- *Review (L).* Review articles, such as literature reviews or chapters in textbooks/reference books. May contain detailed accounts of reported studies but may also be more general in nature.
- *Other articles (O).* This category includes case reports, letters, editorial comments, etc. and summaries from scientific meetings. Animal experiments and technical reports also fall within this category.

In addition to classifying the reports by category, they were also graded for scientific value based on a three-level scale: high, moderate or low. Table 1 provides examples of typical characteristics of these levels for each type of study.

The study category, weight of evidence and number of patients in the study are detailed for each reference in the specific chapters. These characteristics for all references are also summarised at the end of each chapter in tables, as exemplified in Table 2. Table 2 presents all references that were classified and graded for the present report. In total 1496 scientific reports involving 558743 patients were, thus, classified and graded in this report.

Several problems were faced in the assessment of the scientific literature. One was the difficulty to comprehensively cover each field only based on literature searches in databases as discussed above. Another problem relates to the literature classification. Whereas the classification with respect to type of report is mostly straightforward, the quality assessment is amenable to subjectivity. In the light of the vast number of reports to be classified, no effort was made to quality assure this assessment. However, individual examples brought up for discussion did not indicate major differences between the authors in the quality assessment.

A third problem was to detail the number of patients included in the individual reports. This was due to complex trial designs, lack of complete information as well as the use of partially identical cohorts of patients in several reports. Thus, while the correctness of the exact number of patients in individual reports cannot be guaranteed, it is considered more important to indicate the approximate size of the patient cohorts on which the conclusions are drawn.

MAKING INFERENCES BASED ON PUBLISHED DATA

There are models available for conformal formulation of grades of evidence and strength of recommendations based on the quality of the evidence in the literature (7). In the present report, these principles were applied for the diagnosis specific chapters, although the conclusions for each chapter are not presented in such tabular format but rather as individually formulated plain text. In Summary and Conclusions (8, this issue) and in the prospective survey (9, this issue), in which the use of chemotherapy is put in relation to the evidence in the literature, the principles for the formulation of recommendations as outlined

Table 1

Basis for classifying the literature

Type of study	Scientific quality = weight of evidence	
	1 = High	2 = Moderate 3 = Low
M	Thorough reanalysis of well-defined original data from all studies	Summary based on a few studies and/or solely on reported findings.
C	Large, well-monitored multicenter studies that adequately describe protocols, patients, and methods including treatment methods. Patient data sufficiently large to address the issue.	Randomized studies with too few patients and/or too many arms, yielding insufficient statistical power. Poor accounting of the data, high drop-out. Poor description of treatment methods.
P	Well-defined issue, sufficient patient data, adequate statistical methods	So-called 'quick-and-dirty' studies.
R	Large, consecutive patient data that are well defined and analyzed by adequate statistical methodology (e.g., multivariate analysis, case-control methods, etc.). Long follow-up period.	Limited patient data, insufficiently defined, inadequate follow-up time, inadequate statistical methods.
L	Thorough literature review, well-documented patient data often presented in table form.	Reports with incomplete reference citations and poorly supported conclusions.
O	High-quality textbook chapters.	

Table 2

The studies that appear in the reference lists of the diagnosis specific chapters (this issue) were classified and graded as follows (number of studies/number of patients)

Type of study	1 = High	2 = Moderate	3 = Low	Total
M	27/142 118	11/52 546	1/18 000	39/212 664
C	309/125 664	267/47 698	220/43 193	796/216 555
P	109/15 709	194/11 490	133/6 120	436/33 319
R	63/42 203	45/20 944	14/10 563	122/73 710
L	44/15 048	18/7 447	1/—	63/22 495
O	32/—	8/—	—/—	40/—
Total	584/340 742	543/140 125	369/77 876	1 496/558 743

M = meta-analysis, C = controlled clinical trial, P = prospective trial, R = retrospective study, L = literature review and O = other studies.

by the American Society of Clinical Oncology are described and applied (10). One obvious problem in the formulation of conclusions and recommendations when numerous reports of differing evidence quality are available, is how to transmit the quality assessment of a study into a weight in making the final conclusions, not the least when there are, as is often the case, contradictory findings. Although there are attempts to formalise the impact of individual studies in the settling of the final conclusion, this step still includes assessments prone to be subjective. Since all literature on which the conclusions are based on is detailed with respect to type of study and quality, it is in principle possible for the interested reader to make his/her own assessment.

Presentation and interpretation of clinical trial data in the form of meta-analyses have become increasingly popular and data from such analyses are frequently referred to in several of the diagnosis-specific chapters. Furthermore, they often form the basis for the formulation of conclusions in these chapters, not only with respect to the quality of a treatment effect but also for its quantification. Although meta-analysis is a powerful technique, there might be discrepancies between meta-analyses and high-quality large prospective randomised trials, both with respect to direction of treatment effects and, more so, their size (11, 12). Thus, also meta-analyses should be interpreted with caution and the final conclusions for a therapeutic field should preferably also be based on judicious conclusions of results from individual well-performed controlled clinical trials.

ENDPOINTS USED IN CANCER CHEMOTHERAPY TRIALS

A clinical trial endpoint could be defined as to what could be measured in assisting the finding of answers to the trial goals. With respect to efficacy, the endpoint(s) should reflect, directly or indirectly, meaningful benefit at the level of the patient (13). A number of efficacy endpoints are in use and effects can often be expressed in various ways. Comparisons between reports might, thus, be difficult and different ways to express, e.g. survival benefit, within and between chapters had to be accepted. Since the understanding of the most commonly used endpoints in this report are crucial, they are defined below and are followed by a critical appraisal of their meaning. Observe that although these endpoints are often referred to as 'standard', various definitions are sometimes used (14). Thus, each study report should ideally present the endpoint definitions used.

- *Tumour response rate*—the proportion of patients whose tumours respond to the chemotherapy, i.e. decrease significantly in tumour volume (see further below). Confirmation of the response at a second assessment at least 4 weeks later is currently mostly considered mandatory. Response evaluation should preferably be objective, i.e. in essence rely on radiological methods. Assessment of tumour response rate by reporting of changes in tumour markers alone is not yet generally accepted.
- *Time to tumour response*—time from start of treatment to observed objective response.
- *Response duration*—time from start of treatment or observation of response to the time point of observed recurrence or progression.
- *Progression-free survival*—time from start of treatment to the time point of observed disease progression. Equivalent to time to progression. On a population basis, this endpoint is often expressed as the median progression-free survival and/or as the proportion progression-free at a specific time point.
- *Disease-free survival*—for adjuvant treatment time from start of treatment to relapse. On a population basis, this endpoint is often expressed as the median disease-free survival and/or as the proportion disease-free at a specific time point. In advanced disease time from disappearance of the tumour to recurrence
- *Overall survival*—time from start of treatment to death. On a population basis, this endpoint is often expressed as the median overall survival and/or as the proportion alive at a specific time point.

- *Symptomatic improvement/stabilisation*—e.g. the proportion of patients with treatment-related improvement in defined symptoms or the time to occurrence or worsening of a disease-related symptom. This endpoint can be measured and reported in a number of ways.
- *QoL benefit*—description by patient, as assessed by various types of questionnaires, of treatment-related improvement in factors supposed to reflect QoL. Mostly reported as changes over time in various domain scale scores.
- *Tolerability*—the incidence and severity by patient and/or treatment cycle of adverse events, with specification of those considered serious and with severity indicated according to an established toxicity scale. Further detailing of important adverse events with respect to time-points for occurrence, duration, reversibility, appearance, dose-dependency, treatment, etc., provides complementary information. The long-term incidence of secondary malignant tumours, other treatment-related disease, as well as impaired functional conditions, might also be included although these aspects are more frequently addressed in surveys specifically addressing long-term adverse effects.

Some of these endpoints, e.g. tumour response rate and progression-free survival, could be considered as surrogate endpoints, i.e. an endpoint that is more conveniently assessed than a true endpoint, which, however, it is supposed to reflect (see further below).

Considering the number of endpoints in use and also the number of statistical principles for data presentation, it is evident that trial results can be presented in many ways. However, tolerability is almost always included and mostly presented along the lines detailed above. The most commonly used principles for the reporting of efficacy are exemplified below.

In the adjuvant setting, long-term disease-free and overall survival are the most important efficacy endpoints. They are often expressed as the proportion of patients in different treatment arms being disease-free and alive at a specific time-point, e.g. five years post-treatment. When comparing two treatments, the complete survival curves for survival over time should also be considered and differences reported, e.g. by log-rank statistics and hazard ratio including its confidence interval. The long-term benefit of one over another treatment arm might be expressed as the absolute or relative treatment effect. Consider a trial in early breast cancer with an adjuvant arm and a no-treatment control arm and with 1000 patients in each arm. If the overall 5-year survival in the treatment arm is 75% and that in the control arm is 60%, the absolute survival gain is 15%, i.e. 150 patients, whereas the relative treatment effect can either be expressed as a 25% (150/600) increase in survival or a 37.5% (150/400) reduction in mortality. The number of patients needed to treat, unfortunately a seldomly used way to report data, in order to save 1 patient at 5 years would be 6.7 (1/absolute risk reduction). It could also be said that 600 patients in the treatment arm are treated unnecessarily since they were cured by surgery alone.

In the palliative setting, tumour response rate is still a frequently used and main endpoint. However, the reporting of differences in progression-free and overall survival, expressed as gain in median survival and hazard ratios based on the survival curves, is considered to more accurately mirror patient benefit. Furthermore, differences in QoL scale scores over time are frequently reported whereas symptomatic benefit might be reported as the proportion of patients achieving a certain level of symptomatic relief or, alternatively, as the median time to occurrence or worsening of specific symptoms.

In the preoperative setting, finally, the proportion of patients converted to becoming operable as well as the proportion of patients possible to operate with curative intent, i.e. apparent removal of all tumour tissue, are commonly used endpoints, together with disease-free and overall survival.

A CRITICAL APPRAISAL OF THE ENDPOINTS USED IN CANCER TRIALS

A significant tumour volume decrease has long been defined as at least a 50% decrease in the sum of 'volumes' of the index tumour lesions and is mostly assessed objectively by radiological methods. Tumour response is normally separated into partial response (PR), i.e. significant but not complete disappearance of all tumours and their complete disappearance which is denoted complete response (CR). Although tumour response criteria are often referred to in study reports as 'standard', they need to be further defined, e.g. according to WHO (15), and with respect to, e.g. principles for selection of index lesions and requirements for duration of response. Still, there is variation between study protocols on the response definitions and procedures for response assessments. There is also a variation with time in that the number of responses are mostly lower in modern than in previous trials. Reasons for this include more precise radiologic assessment methods, e.g. computerised tomography rather than plain x-ray or scintigraphy, requirements for confirmation of the response and, particularly, independent reviewing of the documentation behind the response evaluation.

The principles for the assessment of tumour response are presently under re-evaluation within a working group recruited from the US and Canada National Cancer Institutes and the European Organisation on the Research and Treatment of Cancer (16). In essence, uni-dimensional assessment of tumour 'volume', i.e. measurement of the longest diameter only, is suggested to replace the current

bi-dimensional measurement, i.e. calculation of the product of the longest diameter and its perpendicular diameter.

In haematological malignancies, it is now possible, by use of molecular biology techniques, to define a state of molecular complete remission in which expression of, e.g. a disease specific gene change, is no longer detected. The importance in routine care of this detailing of the antitumour efficacy is not yet clear.

Although in some cases a tumour response might be directly beneficial to the patient, e.g. by the relief of volume-related pressure on critical tissues, tumour response could not be regarded as a fully justified surrogate endpoint for clinical benefit (13, 17). In, e.g., advanced non-small cell lung cancer, the relationship between response rate and survival in randomised trials seems to be extremely poor. In advanced colorectal cancer this relationship is highly statistically significant but the survival gain for a defined increase in response rate is very small (18). In general, this relationship has not been extensively investigated. From a biological point of view, a poor relationship between tumour response and survival does not necessarily contradict a relationship between tumour response and beneficial symptomatic effects. In fact, there seems to be a consensus among experienced oncologists that there is a positive and strong relationship between tumour response and patient benefit in terms of, e.g. symptom relief and general performance. Furthermore, such palliation has been reported even in the absence of objective tumour remission (19). Several other examples of this are given in the separate diagnostic chapters, particularly about colorectal cancer (20) and pancreatic cancer (21). Although these observations are seemingly correct, there are methodological problems in the reporting of these palliative effects. Thus, from a formal methodological point of view, this issue is also in need of further evaluation prior to the formulation of general evidence-based conclusions (22). Biomarkers have not been established as good surrogate markers either for survival or for symptomatic improvement (22). Thus, tumour response should at first hand be regarded rather as an indicator of antitumour activity that might be worthwhile to study further in comparative trials using endpoints more directly measuring patient benefit.

With respect to response rates, there might also be considerable differences in the way investigators collect and report the data and this contributes to the wide ranges of efficacy reported for essentially identical treatments (14). Although the definitions of PR and CR are reasonably standardised, a number of other factors might affect the response rate reported (14). Furthermore, the way 'objective' tumour volumes are assessed by radiological methods might clearly affect the observed response rate (23).

When adding to this the questionable surrogate value of tumour response rate for true clinical benefit, it is evident that phase II data mostly cannot provide efficacy data making it possible to define the clinical value of a treatment regimen (13). An exception to this rule might be when a new treatment is overwhelmingly effective compared with historical experience. In this case, it has to be common knowledge that the disease status cannot be improved by existing therapy (22, 24).

Time to response, response duration, progression-free and disease-free survival are dependent on the intensity of patient surveillance making between-study comparisons difficult. Time to progression is mostly defined by objective, often radiological, measurements of tumour size, and a prolongation of time to progression is often perceived to reflect clinical benefit. Although this is probably true in most cases, it is not necessarily true for all tumour types or tumour locations and, thus, needs to be individually justified for the various settings studied. In the case of clinical trials in tumour types and settings in which active next-line chemotherapy is available, time to progression has often been given more weight since overall survival might become biased by therapy after the trial.

Whereas disease-free survival reflects the 'curative' potential of a treatment, that of overall survival includes death due to any cause, thus including also possible treatment-related serious adverse effects. Overall survival might be affected by additional therapy provided after observed relapse, thereby distorting the conclusions regarding a survival difference induced by initial treatment arms. In general, progression-free, disease-free and overall survival are considered as 'strong' endpoints with respect to reflection of patient benefit. However, in, e.g. indolent lymphomas, tumour response is considered, based on extensive clinical experience, strongly associated to patient benefit, in terms of symptom relief. Symptomatic improvement is conceptually simple and is reasonably an important endpoint in palliative cancer treatment trials. However, this endpoint is used surprisingly infrequently in clinical trials. To be informative, the type of symptoms and their degree of improvement considered significant need to be justified, which in turn places requirements on the assessment scale and its validity in reflecting clinically significant changes. Symptom scales are mostly included in the QoL questionnaires. The 'objective' assessment of changes in symptoms by, e.g. the physician or a research nurse, is sometimes used (25). However, as for patient reporting, the risk for bias due to knowledge on treatment arm needs to be considered. External 'blinded' evaluation might overcome this problem but is seldomly used.

A number of 'validated' QoL questionnaires are available and QoL assessments using such scales are increasingly included in cancer clinical trials. This is principally supported since for large patient groups chemotherapy will hardly lead to major improvements in survival, making the

issue of therapy-induced improvements in QoL highly important. However, although it is easy to intuitively understand the meaning of QoL, its measurement still suffers from conceptual and methodological difficulties. Thus, it is difficult to define what truly reflects QoL. Furthermore, also in well-controlled and executed clinical trials, assessment of QoL mostly suffers from a number of methodological problems, including question multiplicity, open trial design, lack of pre-defined hypotheses on differences anticipated, difficulties in assessment of clinical relevance, and attrition of data over time in study. These issues are discussed in more detail in the chapter on QoL (3). Thus, although QoL is intuitively a 'strong' endpoint seemingly directly related to clinical efficacy, the information obtained by QoL assessments unfortunately mostly does not allow unambiguous conclusions on clinical benefit. This area is clearly in need of further methodological and conceptual development to provide clinically useful information.

A surrogate endpoint is easier to assess than the endpoint being most clinically relevant, is a prognostic factor, and differences in the surrogate endpoint between treatment groups should translate into differences in the endpoint of interest. Furthermore, ideally it should be possible to estimate the effect on the endpoint of interest from the effect on the surrogate endpoint (26, 27). Surrogate endpoints are commonly used in clinical trials, e.g. blood pressure in cardiovascular diseases, ocular pressure in glaucoma and CD4 counts in HIV research. In oncology, tumour remission rate and changes in tumour markers such as PSA and CA-125 are examples of surrogate endpoints. Unfortunately, very few surrogate endpoints used in medical research, including oncology, have been satisfactorily validated.

In conclusion, the reporting of clinical benefit from chemotherapy in cancer should not rely on a single endpoint. Rather, the assessment of clinical benefit needs to be based on the total outcome based on several endpoints. In treatment settings with a clear curative potential, disease-free and overall survival are the most important whereas in the palliative setting treatment-related survival prolongation, extension of progression-free survival, symptomatic improvement and QoL aspects often all need to be considered and put in relation to the treatment-related toxicity and inconvenience in the final judgement of clinical benefit at the patient cohort level. In some cases individual patients may have strong preferences regarding treatment strategy. This may then have a stronger impact on choice of treatment than recommendations based on the scientific literature.

SOME BASIC ASPECTS OF CANCER CHEMOTHERAPY TRIAL DESIGN AND INTERPRETATION

Essentially all confirmatory trials in oncology, i.e. trials providing key evidence possible to use as a basis for treatment recommendations, need to be comparative, i.e. two or more treatment alternatives are compared in a parallel group, prospective and randomised design. However, in contrast to many other trials investigating the efficacy of pharmaceuticals, chemotherapy trials are essentially never blinded, i.e. both the caregivers and the patients will be aware of the assigned treatment.

Although the open design seems less probable to affect the 'objective' endpoints, such as progression-free, disease-free and overall survival, bias cannot be completely excluded. Thus, the open design might, e.g. induce an unconscious tendency to withdraw patients earlier from a treatment arm using an 'old' compared with a 'new' drug or regimen and this might affect clinical outcome. With respect to 'subjective' endpoints, i.e. symptomatic improvement and QoL, the open design might induce bias irrespective of whether the endpoint quantification is done by the caregiver or by the patient.

Phase II, non-comparative trials in oncology can very seldom provide data allowing an assessment of the true clinical value of a new drug or regimen, not even when well-performed and including many patients. The reason for this is that the efficacy data that can be extracted from such trials will depend on the type of patients recruited to the study. This patient selection factor is hard to quantify and might often not be fully apparent even from a detailed scrutiny of the trial inclusion/exclusion criteria and descriptions of the study population characteristics. Patient selection is probably an important factor behind the often dramatically different results in, e.g. tumour response rate, observed in different phase II trials using essentially identical treatment and patient eligibility criteria. Thus, even very promising findings in phase II trials mostly need to be confirmed in comparative phase III trials (28).

In principle, given the mostly modest improvement in treatment outcome to be expected from 'new' compared with 'standard' chemotherapy for cancer, only the prospective randomised comparative trial design can provide data allowing reasonably safe conclusions on efficacy and safety. Furthermore, these trials mostly need to be large, necessitating multi-centre and multi-national participation.

Key data on which to base routine treatment decisions, will, therefore, derive from patient populations not necessarily fully representative for Swedish conditions. Another potential disadvantage with large trials is that major patient groups for considerable time might be unavailable for smaller exploratory phase II trials that theoretically could indicate major progress in a treatment field. On the other hand, the multi-centre and multi-national approach proba-

bly introduce random bias that tends to decrease differences between treatments. Thus, if a significant efficacy difference is observed in a large trial, this is probably a robust finding.

Patient selection is also of greatest concern in retrospective analyses of a patient material collected at a single or multiple institutions. Retrospective analyses are also often hampered by variable quality of collected data, and the results of these studies can generally only be used for hypothesis-generation. This, e.g. refers to the often large international data bases of patients treated for various malignancies with high-dose therapy followed by stem cell support. Case-control matching within the cohorts may, however, improve the possibilities to define fully comparable patient groups, and supportive evidence to randomised trials may be achieved.

Truly population-based patient materials will diminish the problem with patient selection considerably. Thus, even if difficulties still may be present in the interpretation of whether a change in outcome will be the result of a particular treatment, other interventions or the effects of other confounding factors, these studies may provide important knowledge. If the studies use a prospective protocol to follow, complete registration of all incident case and proper statistical analyses, the value may not be far behind that of randomised trials. The possibilities for such studies are particularly prominent in the Nordic countries. In rare tumours, or unusual presentations also of common tumours, properly designed, observational studies may be the only way to improve knowledge.

Unfortunately, efficacy of cancer chemotherapy efficacy is still often claimed on the basis of non-controlled phase II trials reporting tumour response rate. However, even in the presence of high-quality phase III trials there will still be a problem regarding the generalisability of the trial results, i.e. the extent to which an observed efficacy level can actually be achieved when using the treatment in a routine clinical setting. A reduction in expected benefit will most probably occur if the new treatment is used in patients with a poorer prognosis, e.g. due to age, more advanced tumour stage, impaired performance status, or more extensive prior chemotherapy. This becomes more critical the smaller the treatment effect is; to observe a chemotherapy induced gain in median survival in the range 1–3 months, as in advanced non-small cell lung cancer, the therapy might have to be restricted to patients with favourable prognostic factors. This should, ideally, be taken into account in the formulation of treatment recommendations by expert panels, i.e. not only should the type of patients who should receive the treatment be identified, but also the type of patients who should not.

SOME COMMENTS ON THE DIAGNOSIS-SPECIFIC CHAPTERS

It is clear from the literature that treatment recommendations in cancer chemotherapy are often based on clear evidence from the literature. However, in many areas, also in those encompassing many patients and in highly resource-consuming treatment settings, surprisingly little controlled confirmatory data are available. The lack of good data on the impact of chemotherapy on survival and QoL in metastatic breast cancer as well as the benefit from stem-cell supported high-dose chemotherapy in most tumour types are examples. In those situations, when facing the individual patient, the physician in charge needs to give treatment advice that will necessarily be influenced by more or less well-founded subjective preferences. It was initially sometimes possible to track such influence also in the formulation of the chapter summary and conclusions. However, the conclusions were subsequently refined within the project group with the requirement that they should be based on studies with sufficient quality and size.

It is also clear that the quality of the data differs considerably between diagnoses but also that 'traditions' might influence data interpretation. Thus, in adjuvant treatment of breast cancer and and in treatment of metastatic non-small cell lung cancer there are now a number of high-quality randomised trials available whereas in indolent lymphoma and chronic lymphocytic leukaemia conclusions often have to be formulated based on phase II data. Therapeutic 'nihilism' in a diagnosis might produce a more conservative assessment than in a traditionally more treatment prone area, as indicated when comparing, e.g. pancreatic cancer with chronic lymphocytic leukaemia. Again, such differences between chapters are difficult to avoid but were tried to be considered in the formulation of the overall conclusions.

Furthermore, in the individual chapters the conclusions are often quite general and do not explicitly consider the problem of extrapolating clinical benefit as observed in clinical trials to the clinical routine situation as discussed above. Finally, the reporting of QoL data in the diagnosis-specific chapters in most cases follows the interpretation made in the published study reports. However, further discussions within the project group led to a more critical appraisal of the methodology of QoL assessment, as evident from the specific chapter on assessment of QoL (3). Thus, the QoL sections of the diagnosis specific chapters are suggested to be read in light of the methodological difficulties as discussed in the specific QoL chapter.

CHEMOTHERAPY FOR CANCER IN A FUTURE PERSPECTIVE

Some future perspectives of the development of cytotoxic drugs for use in cancer treatment were discussed in 'What is cancer chemotherapy' ((29), this issue). Only some more general aspects will be addressed here.

The knowledge at genetic level of factors explaining the cause and progress of cancer has 'exploded' during the last decade and as a consequence there is much hope that this knowledge will also open for completely new treatments, directly targeting various defects at the gene level. This field, i.e. gene therapy, is already in its initial clinical phase although it is still too early to judge whether these treatment approaches will revolutionise cancer treatment. Until this is known, the presently used 'non-specific' cytotoxic chemotherapy will remain a mainstay of treatment of advanced cancer and in several adjuvant settings. Furthermore, it might also be that cytotoxic chemotherapy will remain an important part of cancer treatment, as part of combined treatments, also if gene therapy and other approaches prove to be useable.

Besides gene therapy, a highly interesting emerging field that could be broadly included under the heading chemotherapy is the use of the broad array of substances supposed to inhibit angiogenesis, inhibit development of metastases or slow tumour growth. These agents are now in early to late clinical development trials. Drugs based on these approaches will probably not have the potential to eradicate tumours but might rather induce 'tumour dormancy' which means a conceptual change whereby cancer might turn into a chronic disease compatible with long-term survival. This development will certainly require a change in our view of cancer and will also require new trial designs considering definition of suitable dosing schedules as well as for assessment of efficacy (13). Also in this case, the combination of these drugs with cytotoxic chemotherapy, to achieve tumour volume reduction, might become the most suitable approach.

Another interesting development is non-ablative or so called 'mini'- transplantations in which chemotherapy is combined with administration of haematopoetic stem cells from a related donor with the aim to obtain an immunological reaction towards the tumour cells.

Thus, even in light of the ongoing highly interesting development and given also the often difficult and slow path from the proving of a pharmacodynamic principle to a standard treatment, it seems reasonable that cytotoxic chemotherapy will remain a necessary ingredient in the treatment of cancer for quite some time.

REFERENCES

1. Medical Research Council. Sjukvårdens prioriterings- och rationaliseringsinstitut. Chemotherapy in cancer. In: Glimelius B, Hafström Lo, eds. Consensus conference. Stockholm 1990, pp 146. (in Swedish)
2. Glimelius B, Brorsson B. Chemotherapy. Changed practice due to new research – not a consequence of a consensus statement. Läkartidningen 1993: 1937–41. (in Swedish)
3. Gunnars B, Nygren P, Glimelius B. Assessment of quality of life during chemotherapy. Acta Oncol 2001; 40: 175–84.
4. Karlsson G. Economic aspects of chemotherapy. Acta Oncol 2001; 40: 412–33.
5. Deeks JJ. Systematic reviews of published evidence: Miracles or minefields. Ann Oncol 1998; 9: 703–9.
6. Haward RA. Preparing guidelines and documented clinical policies. Ann Oncol 1998; 9: 1073–8.
7. Goodman C. Literature searching and evidence interpretation for assessing health care practices. Swedish Council on Technology Assessment in Health Care. Norstedts, Stockholm 1993, pp 81.
8. Glimelius B, Bergh J, Brandt L, et al. The Swedish Council on Technology Assessment in Health Care (SBU) systemic overview of chemotherapy effects in some major tumour types. Summary and conclusions. Acta Oncol 2001; 40: 135–54.
9. Ragnhammar P, Brorsson B, Nygren P, Glimelius B. A prospective study of the use of chemotherapy in Sweden and assessment of the use in relation to scientific evidence. Acta Oncol 2001; 40: 391–411.
10. Desch CE, Benson AB III, Smith TJ, et al. Recommended colorectal cancer surveillance guidelines by the American Society of Clinical Oncology. J Clin Oncol 1999; 17: 1312–21.
11. LeLorier J, Grégoire G, Behhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 1997; 337: 536–42.
12. Villar J, Carroli G, Belizán JM. Predictive ability of meta-analyses of randomised controlled trials. The Lancet 1995; 345: 772–6.
13. Eisenhauer EA. Phase I and II trials of novel anti-cancer agents: Endpoints, efficacy and existentialism. Ann Oncol 1998; 9: 1047–52.
14. Tonkin K, Tritchler D, Tannock I. Criteria of tumour response used in clinical trials of chemotherapy. J Clin Oncol 1985; 3: 870–5.
15. Miller AB, Hoogstraten B, Staquet M, et al. Reporting results of cancer treatment. Cancer 1981; 47: 207–14.
16. American Society of Clinical Oncology. New international unidimensional response criteria for solid tumours: simpler and standardised. 35th annual meeting; special session 1999.
17. Buyse M, Piedbois P. On the relationship between response to treatment and survival time. Stat Med 1996; 15: 2797–812.
18. Buyse M, Thirion P, Carlson RW, et al. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Lancet 2000; 356: 373–8.
19. Constenla DO, Hill ME, A'Hern RP, et al. Chemotherapy for symptom control in recurrent squamous cell carcinoma of the head and neck. Ann Oncol 1997; 8: 445–9.
20. Ragnhammar P, Hafström Lo, Nygren P, Glimelius B. A systematic overview of chemotherapy effects in colorectal cancer. Acta Oncol 2001; 40: 282–308.
21. Permert J, Hafström Lo, Nygren P, Glimelius B. A systematic overview of chemotherapy effects in pancreatic cancer. Acta Oncol 2001; 40: 361–70.
22. American Society of Clinical Oncology. Outcome for technology assessments and guideline development. www.asco.org/prof/pp/html/m_outcome3.htm 1999.
23. Gwyther S, Bolis G, Gore M, et al. Experience with independent radiological review during a topotecan trial in ovarian cancer. Ann Oncol 1997; 8: 463–8.
24. Staquet M, Dalesio O. Designs for phase III trials. In: Buyse ME, Staquet M, Sylvester RJ, eds. Cancer clinical trials. Oxford: Oxford University Press, 1992: 261–75.
25. Sprangers M, Sneeuw KC. Are health care providers adequate raters of patients' quality of life – perhaps more than we think? Acta Oncol 2000; 39: 5–8.
26. Jonsson B. Problems relating to difficult medical areas, eg cancer and AIDS drugs; Preclinical and Clinical Evaluation of

Medicinal Products. Nordiska Läkemedelsnämnden 1996; 41: 47–55.
27. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomised experiments. Biometrics 1998; 54: 1014–29.
28. O'Connell MJ. Etoposide, doxorubicin, and cisplatin chemotherapy for advanced gastric cancer: an old lesson revisited. J Clin Oncol 1992; 10: 515–6.
29. Nygren P. What is cancer chemotherapy. Acta Oncol 2001; 40: 166–74.