From clinical trials to evidence-based medicine: how to build the evidence!

Patrick Therasse

EORTC Data Center, Av. E. Mounier 83/11, 1200 Brussels, Belgium

Introduction

Until recently, most therapeutic advances have resulted from incidental observation and hypothesis such as the discovery of penicillin by Fleming, which as a treatment for pneumonia (then a life-threatening disease) did not require confirmation with a randomised controlled experiment. These are rare cases in the history of medicine, as most diseases have multifactorial physio-pathological characteristics, and as it is exceptional to find simple treatments which can easily cure life-threatening diseases.

Malignant tumours are characterised by their diversity, heterogeneity and phenotypic variability. In a field as large and complex as oncology, improvement of knowledge invariably necessitates the discovery of new techniques and treatments on the basis of homogeneous and reliable data gathered from experiments far beyond the basic clinical observation. Claude Bernard was a pioneer in the development of objective and rigorous medical experimentation, making observations on a group of patients rather than single case reports. In fact, clinical trials represented a new step in the development of medicine, supporting advances in clinical science by proven facts. A new era has been opened for scientists and clinicians who have now at their disposal a tool enabling them to provide convincing scientific arguments based on solid experimentation to demonstrate the efficacy of a new therapy.

Nowadays, our knowledge of cancer treatment and cancer biology evolves rapidly with more than 3000 publications related to cancer indexed in Medline every month, out of which about 100 are randomised clinical trials and 200 are non-randomised clinical trials [1]. The need to differentiate the good from the bad and useful from useless information has become obvious. In this perspective, A. Cochrane with his concept of "evidence-based medicine" was the first scientist to highlight the necessity to support changes in clinical practice with proven data. Clinical trials were therefore classified on the basis of their actual contribution in establishing the scientific evidence following which therapeutic decisions should be taken.

The main purpose of this chapter is to review the role and the importance of randomised phase III trials in contributing to the establishment of "evidence" for the development of new therapeutics in oncology. For doing so, the standard pathway of therapeutic development in oncology and evidence-based medicine are reviewed. A larger section is then dedicated to randomised controlled trials illustrating with examples issues of size, replication and randomisation. The last two sections cover the roles of meta-analyses and the development of clinical practice guidelines.

Development of a new anticancer treatment

In oncology, the development of new treatments has usually followed a fairly standardised pathway. New anticancer agents with promising preclinical characteristics are introduced in human patients in phase I clinical trials to identify the dose and schedule under which the agent should be subsequently investigated. Historically, the primary endpoint of a phase I clinical trial has been the determination of the maximum tolerated dose (MTD) and the dose-limiting toxicities (DLT). This concept was based on the empirical observation of a dose–response relationship between doses of cytotoxics and tumour shrinkage. MTD is used to determine the dose to be administered at the next stage of development of the anticancer agent. Nowadays, new classes of anticancer agents (non-cytotoxic) may be active at doses which do not generate dose-limiting toxicities and the dose at which a maximal biological activity is recorded (also known as the optimal biological dose) may be a preferred endpoint. New combinations of existing anticancer agents may also be investigated for their respective dose and schedule in the combination during phase I clinical trials. New radiotherapy

techniques or new radio-chemotherapy combinations are also investigated in phase I trials with acute toxicity and feasibility endpoints.

Once a dose and a schedule have been identified, phase II clinical trials are conducted which aim at identifying and quantifying the level of anticancer activity. The dose and schedule of the new anticancer agent will also be reviewed on the basis of the toxicity data accumulated. The design of the phase II studies can vary according to stage of development of the new treatment (early versus late) and the main objective of the study (activity versus toxicity versus feasibility). Randomised phase II trials can be set up to test simultaneously several new drugs or to test the combination of treatments using the standard of care as control arm. The purpose is not to compare treatment arms but to use the randomisation to avoid selection biases and confirm in real life the assumptions on the basis of which the trial was developed.

Once a new treatment has been demonstrated to be potentially more promising (more active or less toxic) than the standard of care in a particular indication, it will be tested in randomised phase III trials to demonstrate (unequivocally) the clinical benefit (better efficacy or lower toxicity) brought by the new treatment. The cornerstones of randomised phase III trials have been described by Buyse in another section and will not be discussed here. Under some specific circumstances, this last step can be sufficient to enable the marketing authorisation of new anticancer agents or the modification of standards of care with a new therapeutic strategy. However, in most instances, because the size of the benefit that can be expected (so far) with new anticancer agents or new treatment strategies is small, a second randomised trial or a meta-analysis of several trials will be conducted to confirm the initial findings and support a change in medical practice. Results of phase III trials will also be integrated into consensus conferences and treatment guidelines, which will be contemplated by clinicians to decide on the appropriate treatment for a patient seen in consultation.

The diagram shown in Fig. 1 summarises the development pathway described above. Of course, there can be a lot of variation and alternatives to this scheme, specifically at the phase I and phase II levels, but that is another story for another paper.

Clinical trials and evidence-based medicine

Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. The individual clinical expertise should be based upon the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice. The best available external clinical evidence should be based on clinically relevant research, often from the basic sciences of medicine, but especially from patient centred clinical research into the efficacy and safety of treatments. External clinical evidence both invalidates previously accepted treatments and replaces them with new ones that are more efficacious and/or safer.

Evidence-based medicine is not restricted to randomised trials and meta-analyses! It involves tracking down the best external evidence with which to answer our clinical questions. To find out about the accuracy of a diagnostic test, we need to find proper cross-sectional studies of patients clinically suspected of harbouring the relevant disorder, not a randomised trial. For a question about prognosis, we need proper follow-up studies of patients assembled at a uniform, early point in the clinical course of their disease. And sometimes the evidence we need will come from the basic sciences such as genetics or immunology. It is when asking questions about therapy that we should try to avoid the non-experimental approaches, since these routinely lead to false positive conclusions about efficacy. Because the randomised trial, and also the systematic review of several randomised trials, is so much more likely to inform us and so much less likely to mislead us, it has become the "gold standard" for judging whether a treatment does more good than harm. However, some questions about therapy do not require randomised trials (successful interventions for otherwise fatal conditions) or cannot wait for the trials to be conducted. And if no randomised trial has been carried out for our patient's predicament, one must follow the trail to the next best external evidence and work from there [2].

Over the last decade, evidence-based medicine has received an increasing amount of attention, and there are many papers and books dedicated to this subject. Also, several definitions of "level of evidence" have been advocated [3–7]. The most widely used classification systems for strength of medical evidence are based on study design. Unfortunately, also in this case, there is no consensus on a single preferable system, but most systems are variations on the theme set out in Table 1. Level I evidence originates only from adequately powered randomised controlled trials, either individual trials or a meta-analysis of trials. While the higher priority given to randomised con-

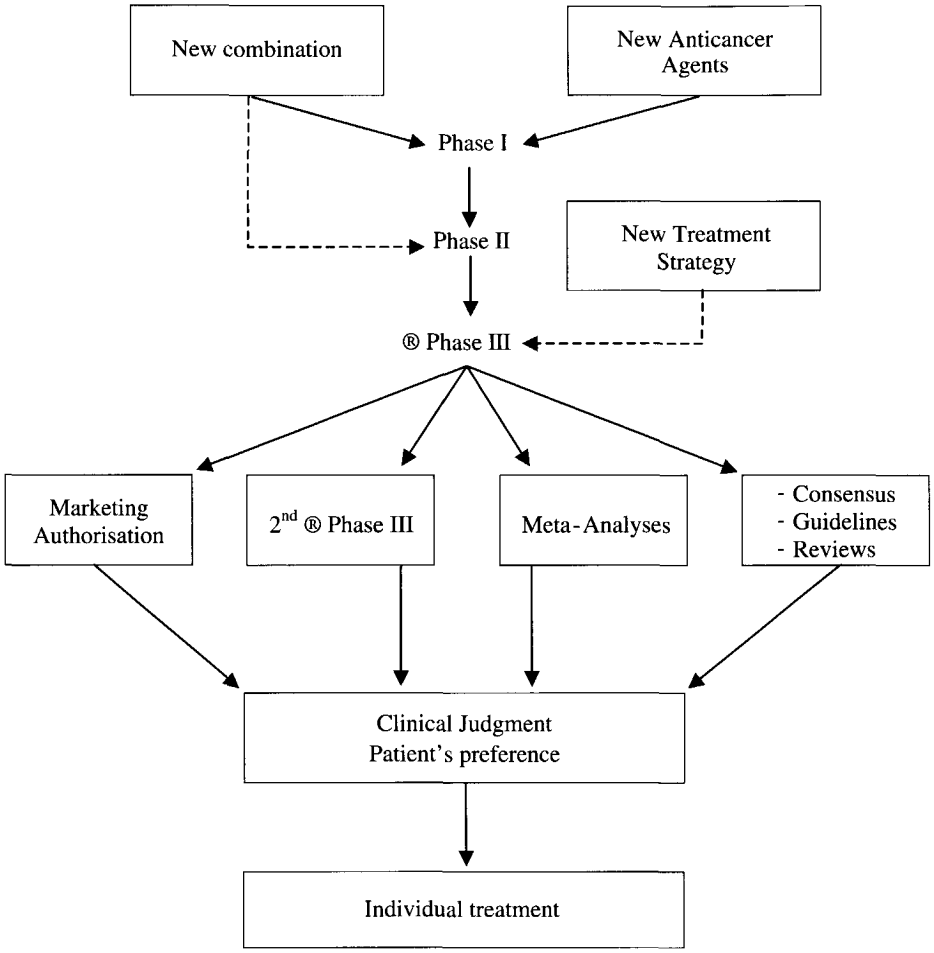

Fig. 1. Development of new therapies in oncology.

Table 1
Simplified scale for classification of evidence based on underlying study design [a]

Level I	Adequately powered, high quality randomised trial, or meta-analysis of randomised trials showing statistically consistent results
Level II	Randomised trials inadequately powered, possibly biased, or showing statistically inconsistent results
Level III	Non-randomised studies with concurrent controls
Level IV	Non-randomised studies with historical controls (i.e. typical single-arm phase II studies)
Level V	Expert committee review, case reports, retrospective studies

[a] Adapted from others, following discussion between Buyse, Bentzen, Tannock and Therasse (May 2003).

trolled trials is fully justified, it is easy to imagine that there are poor randomised (even adequately powered) trials that may be less convincing than well-designed studies at lower levels of evidence.

In 1999, Djulbegovic et al. [8] reviewed the quality of medical evidence in haemato-oncology and identified 783 randomised controlled trials (level I) conducted between 1966 and 1996 pertaining to 24% of common decisions/interventions for haematological malignant disorders. An additional 21% of the decisions/interventions were supported by evidence from single-arm prospective studies (level II). However, only retrospective studies or anecdotal reports (level III) were available to support 55% of the identified decisions/interventions. In a retrospective review of the decisions/interventions made in the management of 255 consecutive patients, 78% of the initial decisions/interventions in the management of newly diagnosed haematologic/oncologic disorders could have been based on level I evidence. However, more than half (52%) of all the decisions/interventions made in the management of these 255 patients were supported only by level II or III evidence. The authors concluded that strong evidence was missing to support about half of the therapeutic decisions taken

Table 2
Randomised controlled trials in the evidence-based oncology database (1975–2001) [9,10]

Tumours completed	Literature review	Retained	Relevant
Head and neck	39	16	7
Oesophagus	184	72	11
Gastric	121	58	14
Pancreas	129	67	6
Colorectal	528	351	21
Sarcoma	32	23	7
Melanoma	295	167	22
Breast	338	300	22
Cervix	48	50	7
Endometrial	25	20	1
Liver	170	93	6
Gall bladder	40	11	1
Small bowel	11	7	2
Anus	7	3	3
Total	1967	1238	130

and also that even when level I evidence was available about 1/3 of the decisions were still based on level II or III evidence.

Brooks et al. [9,10] in the evidence-based oncology project investigated the level I evidence in solid tumours by an in-depth review of the literature published since 1975. They identified 1967 papers published as prospective randomised controlled trials (RCT) across 14 different types of solid tumours (Table 2). After a first simple review, they retained 1238 papers which fit the criteria of randomised trials with clinical endpoints. After a critical review of the selected trials, only 130 (11%) RCT were to be ranked as contributing to the highest level of evidence with a potential impact on clinical practice.

Finally, Cox [7] evaluated in the year 2000 the types of evidence that have fashioned clinical decision making in oncology over the previous century. His work included a review of publications that altered medical practice, a review of the progress made in oncology by looking at the relationship between 5-year survival rates of individual institutions and national mortality rates and an evaluation of the quality of data from clinical trials. Cox reported that "*res ipsa loquitur*" (in other words: the thing speaks for itself) was the dominant mindset of researchers in the first half of the century and continuing into the second half. However, recognition of the scarcity of dramatic improvements in outcome and the possibility of incremental improvements led to the mounting of prospective randomised comparative trials that could identify such incremental improvements. Findings from these trials have profoundly altered patient care over the last 25 years. Importantly, the data collected suggest that there is a sequence of events — increased survival rates in patients at research institutions followed by significant increases in survival rates nationally — followed by a reduction in annual mortality rates that do reflect improvements in treatment independently from other confounding factors such as screening, prevention or health care interventions.

What can we conclude from the review of the literature dedicated to evidence-based-medicine in oncology? First of all, many primary therapeutic interventions in oncology are not supported by strong evidence and, conversely, when strong evidence is available, it is not adequately used. Secondly, there is a lot of clinical research activity being carried out in oncology but very few trials will actually contribute to establish level I evidence. Finally, the recognition that incremental improvements rather than striking improvements are much more likely to happen in cancer treatment has facilitated the development of randomised controlled studies, and the benefit of these trials does translate into an unequivocal reduction of annual mortality rates.

Phase III randomised clinical trials

The randomised controlled clinical trial is the only method known to avoid selection and confounding biases in clinical research. This design approximates the controlled experiment of basic science [11]. The hallmark of randomised controlled trials (RCTs) is assignment of participants to exposures purely by the play of chance. RCTs reduce the likelihood of biases in determination of outcome. It eliminates confounding biases, both known and unknown, tending furthermore to be statistically efficient. If properly designed and done, an RCT is likely to be free of biases and is thus especially useful for the examination of small and moderate effects.

Patient selection and other factors listed in Table 3 prevent data from single-arm trials from giving meaningful information about the relative merit of a new treatment. Such trials are important to show feasibility, to generate evidence of sufficient activity to warrant further testing, and ideally to give information about underlying mechanisms, but cannot disclose evidence of benefit relative to a standard approach [12].

Do we always need randomised controlled trials?

The rationale for RCTs in cancer is based on experience. This experience tells us that in most cancers, major breakthroughs are rare and we may

Table 3
Problems that prevent single-arm studies from giving a definitive answer about therapeutic benefit

Selection of patients	Almost all clinical trials have eligibility criteria leading to the creation of a group of patients that differ from those seen outside of the trial setting. In many instances these patients are in better condition, which may influence the tolerability of the investigational regimen and also the long-term outcome of these patients.
Publication bias	Trials where there appears to be better outcome than prior experience (which is possible by chance alone) tend to be published earlier and more prominently than those that do not.
Stage migration	Stage by stage comparison with a historical control is invalid because newer imaging techniques are more sensitive and tend to upstage patients. Patients in the trial therefore may have less disease than a historical comparison group.
Better supportive care in trials	Patients in clinical trials are often followed more closely and regularly than patients not recruited to trials. Apparent improvements in outcome may be due to better supportive care rather than to the change in anti-cancer therapy. Disease progression can also be identified and treated when there are not yet clinical symptoms of progression.
Better compliance with treatment in clinical trials	Patients recruited to trials may have supportive care adapted to their conditions to enable continuation of the treatment. Patients may be encouraged by their physicians/clinical trial nurse to continue with the protocol treatment.

generally anticipate that even the best new treatments will result in only modest improvements in long-term outcome. Such small differences can only be distinguished from other possible confounding factors in a randomised controlled setting.

The determination of a real treatment benefit extrapolated from small studies (situations referred to as "it speaks for itself") has been restricted to a few instances in oncology such as the treatment of testicular cancer, children's leukaemia, Hodgkin's lymphoma and, more recently, gastro-intestinal stromal tumours. Besides these few examples, the literature is full of small studies where a promising treatment tested in a single-arm trial is found to be no better than standard treatment in subsequent randomised trials. To illustrate this last statement, two concrete and well-known examples are summarised below.

High-dose chemotherapy in breast cancer

In the early 1990s, the preliminary results of phase II trials testing the concept of high-dose chemotherapy with bone marrow transplantation were made public. These results were compared to historical control groups and substantial differences in favour of high-dose chemotherapy emerged after 1 year, with a 40% absolute difference reported after 3 years. The results fuelled the enthusiasm for converting high-dose adjuvant therapy to a standard option for the treatment of high-risk primary breast cancer. Following this promise of a dramatic treatment benefit, more than 41,000 women were treated with high-dose regimens despite the lack of evidence of efficacy. In 1999, two randomised controlled phase III trials comparing the standard of care with high-dose chemotherapy in this population did not demonstrate any tangible benefit of the high-dose approach, which also resulted in a severe increase in morbidity and costs [13,14]. High-dose chemotherapy is still under investigation in randomised trials.

Dose-intensive combinations in non-Hodgkin's lymphoma

In the early 1980s, the results of non-randomised phase II trials investigating five new dose-intensive combinations were published. Sample sizes ranged from 61 to 134 patients, and complete response rates ranging from 61% to 82% were reported, contrasting with the 50% complete response rate historically reported for CHOP (cyclophosphamide, doxorubicin, vincristine, prednisone). Those results generated not only an abundant literature, but also a large number of subsequent studies without any CHOP control arm, aiming at "confirming", "improving" or "comparing" the activity of these new regimens. However, the five major trials that did include randomisation versus the CHOP (or a CHOP-like) regimen, published between 1992 and 1996, unanimously demonstrated that none of the new combinations improved the response rate, progression-free survival or overall survival, and that all resulted in increased toxicity and toxic deaths rates. Miller concluded that "much of the controversy regarding treatment for diffuse large cell lymphoma resulted from inappropriate comparisons of results of new studies to historic CHOP trials without regard for major differences in patients selection" [15]. During nearly 10 years (1983–1992), large numbers of patients were treated with high-dose regimens, inside or outside trials. These patients would have

been spared unnecessary toxic treatment if the initial phase II trials had included randomisation versus CHOP. Should this have happened in the World Wide Web era, patient refusal to participate in randomised phase III trials would even have endangered their completion [14].

How large should the trial be?

As of today, there has not been a systematic review of randomised trials in solid tumours that has shown an absolute long-term survival improvement of > 10%. However, over the last 40 years, many cancer trials have usually targeted differences larger than 15% in long-term survival rates and, therefore, randomising no more than a few hundred patients. The large majority of these trials have not been able to demonstrate any meaningful difference, which could only be demonstrated in some instances by the combination of results from trials in meta-analyses.

On the other hand, very large trials are difficult to set up, requiring the collaboration of a large number of investigators and can be very costly. A careful evaluation of the question that has to be answered and the context under which the trial can be conducted should enable the participants to choose between small and precise trials or large and simple studies [12,16]. Table 4 adapted from Stenning et al. [16] highlights the principal criteria that should be considered to guide the decision in one way or the other.

Do we need to replicate phase III trials?

When a "positive" randomised clinical trial is reported, there is usually a call to perform a second confirmatory trial. There may be debate about whether the first trial is sufficiently convincing on its own and whether further trials are actually needed. In most instances, consensus is unlikely after the publication of a single trial unless the trial is extremely large, and that differences in interpretation of the results are often appropriately based on differences in prior beliefs held about the efficacy of the intervention. These differences in interpretation will lead some investigators to perform a confirmatory randomised trial, whereas others will not require it or be willing to participate in it [17].

Table 5 attempts to delineate all possible factors that may influence the decision to replicate a particular phase III trial. In most instances, it will be a combination of these factors that will lead to the final decision to replicate or not a particular study. Many concrete situations could illustrate the influence of these factors on their own or in combination, but let us limit the examples to two situations where replicated trials have been conducted either in sequence or in parallel.

Paclitaxel/Cisplatin as treatment of advanced ovarian cancer

By mid-1993, the Gynaecologic Oncology Group (GOG) had disclosed the first results of a prospective randomised clinical trial. In this trial, pacli-

Table 4
Two approaches to randomised phase III design

	Small, precise	Large, simple
Context	Is the new treatment safe and promising? Proof of concept principles	Is the treatment of practical use?
Endpoints	Intermediate outcome, such as relapse or progression or even biologically driven hypotheses	Usually overall survival
Data collection	Large amounts to identify protocol deviations and monitor compliance	Minimum required to answer question posed by trial
Population	Homogeneous: the ideal selection criteria to respond to the treatment	Heterogeneous: representative of those to whom the results would be extrapolated
Participants	Those with most experience, most likely to get the "best" out of a new treatment	Representative of those who would use the treatment outside of the trial
Treatments	All aspects of patient care clearly defined	All other aspects of patient care as for normal practice
Surveillance and follow-up	May require specific specialised tests, intensive surveillance	Only investigations used in routine clinical practice
Analysis	May concentrate on those eligible or evaluable (if defined *a priori*)	Analyse all patients

Table 5
Factors possibly affecting the decision to replicate phase III trials

Amplitude of the benefit	A well-conducted trial, with an important benefit (usually beyond initial expectation) demonstrated for a definitive endpoint (relapse or survival) will usually not be replicated. There could be ethical concern or problems of feasibility (recruitment) to reproduce the study.
Feasibility	Positive trials in the setting of rare tumour types for which recruitment was long and difficult are less likely to be reproduced in comparison to identical findings in more frequent tumour types. The feasibility can also be affected by the practical aspects (drug supplies, infrastructure, financing).
Quality of the trial	There may be questions about the quality of the trial in terms of statistical analysis, endpoints, pathology, subgroup analyses, etc. that can by themselves cast doubts on the internal validity of the trial.
Toxicity	Positive trials with marginal clinical benefit and unfavourable toxicity profile (borderline risk/benefit ratio).
Difference in standard of care	Reference treatment arm is not judged to be appropriate or mode and schedule of administration of one or more anti-cancer agents may be different; difference in treatment modalities (including radiotherapy and surgery).
Geographical spread	This is the well-recognised European vs. North American reciprocal lack of trust.
Scepticism of clinicians	It may be sometimes hard to convince clinicians to change practice until they have themselves (or someone they know and recognise) practiced the new regimen.
Registration/marketing initiatives	Sometimes, companies initiated parallel or successive trials with similar designs to speed up the registration process (providing compelling evidence) or to ensure the adequate marketing of a new regimen.
Previous experience in the same setting	If most of the previous trials investigating the same strategies in similar settings have not reported any benefit, it is highly probable that a single trial showing some benefit is a false-positive trial.

taxel (T), combined with cisplatin (P) — a combination denoted as TP — was infused in patients with advanced ovarian cancer for a 24-hour period. This regimen produced a higher response rate and a longer progression-free survival (PFS) in women with newly diagnosed and suboptimal debulked International Federation of Gynaecology and Obstetrics (FIGO) stage III or IV epithelial ovarian cancer than those produced by the "standard" cyclophosphamide-cisplatin (CP) regimen.

A group of European and Canadian investigators found these results to be impressive but not conclusive enough. They believed that (a) further data were required before the TP combination could be adopted as the new standard first-line chemotherapy regimen for this disease, (b) the TP regimen could be improved by increasing the dose of paclitaxel and shortening its infusion time, and (c) more knowledge was needed regarding the comparative quality-of-life and economic impacts of these competing regimens [18].

In 1994, a European–Canadian intergroup initiated a similar trial (with a higher dose of paclitaxel and a shorter infusion time) and recruited 600 eligible patients within 16 months. This level of accrual gave this study an 80% probability of detecting an increase in the median PFS by one third. Accrual of patients in the trial was completed in August 1995, 4 months after GOG publicly reported a highly significant survival advantage in favour of TP and 4 months before these striking results were published in the *New England Journal of Medicine* [19].

The results of the European–Canadian intergroup trial for women with advanced ovarian cancer confirmed the findings of the GOG #111 trial published in 1996 that the combination of cisplatin and paclitaxel confers a survival advantage over the combination of cyclophosphamide and cisplatin. Importantly, they also extend these findings in that the trial included a broader range of patients, was conducted in a largely community-based setting, and included a much higher rate of crossover to paclitaxel on first progression of disease in the standard arm. The fact that the 3-year survival results of this intergroup trial mirrored those of GOG #111 trial had two important implications: (1) it provided strong or level I evidence that the paclitaxel–cisplatin regimen is superior to the cyclophosphamide–cisplatin regimen, establishing this regimen as the gold standard for this disease; and (2) it refuted the claim that administration of paclitaxel should be delayed until relapse since the crossover in the second trial did not influence the results of the trial [18].

However, two other randomised controlled trials (GOG#132 [20] and International Collaborative Ovarian Neoplasma (ICON) 3 [21]) showed no benefit for the addition of paclitaxel to cisplatin/carboplatin regimes. This, together with the observation that there is a myeloprotective effect when paclitaxel

is added to a platinum regime, has led to the notion that there may be a degree of antagonism when these are given concurrently. Both trials concluded that formal randomised studies examining the option of sequential therapy are a logical next step. Sequential therapy allows initial therapy with full doses of the best single agent in ovarian cancer, i.e. carboplatin [22].

Irinotecan/5FU as treatment for advanced colorectal cancer

In March 2000, Douillard et al. [23] reported the results of their study comparing irinotecan combined with fluorouracil (5FU) against fluorouracil alone as first-line treatment for metastatic colorectal cancer. The study was open from May 1997 until February 1998 and recruited 387 patients in 13 European countries. They reported a significant advantage in terms of time to progression and survival for the combination therapy. The reference arm constituted a mixture of 5FU schedules following either the French DeGramont schedule (every 2 weeks) or the German schedule (once weekly) both using an infusional 5FU regimen.

In September 2000, Saltz et al. [24] reported the positive results of their investigation of irinotecan + 5FU/leucovorin over the standard of care (5FU/Leucovorin alone) used in the United States. Advantages were significant both in terms of delaying progression as well as in terms of prolonging survival. The trial recruited over 600 patients in North America and used a standard 5FU bolus perfusion as control arm.

An editorial in the *New England Journal of Medicine* [25] commented on both studies and recognised that this new combination approach could indeed be seen as a new standard of care, pending some improvement in the toxicity profile that could certainly be ameliorated. The company licensing irinotecan presented both trials to the Food and Drug Administration (FDA) to receive marketing authorisation.

Systematic reviews and meta-analysis

A meta-analysis (or overview, or systematic review) is the process whereby the quantitative results of separate, but similar, studies are combined together using formal statistical techniques [26]. It is also discussed by others as an exhaustive, objective, quantitative and systematic review of the best available evidence addressing a specific question (level I evidence) [27].

In most instances, meta-analyses are carried out in situations where successive individual randomised trials produce inconclusive and often conflicting results in a particular clinical situation. They provide an estimate of the overall treatment effect and its precision based on data from all available properly randomised trials. Unfortunately, it also reflects our inability to set up trials which are powerful enough to detect small but clinically relevant differences. Many trials may conclude that there is no difference in treatment efficacy when in fact the results are inconclusive: one important difference may have been missed simply due to a lack of power.

Not all meta-analyses carry the same weight and those conducted on the basis of published trials or summary reports should be interpreted with great precaution. Meta-analyses conducted by pooling the individual patient data of all available randomised trials do not carry the same publication and selection biases as the others and, moreover, offer the possibility to (a) perform some control of the quality of the data provided, (b) update the data sets with recent information, (c) perform subgroup and prognostic factor analysis, (d) perform time to event analyses, estimation of hazard ratios and evaluation of survival curves.

The use of meta-analyses in oncology has become well known thanks to the pioneering efforts of the Early Breast Cancer trialists collaborative group in the late 1980s. This group succeeded in clarifying and quantifying the contribution of surgery, radiotherapy and systemic therapy in the treatment of early breast cancer, with an overview recruiting in total over 100,000 women in more than 200 trials. Since then, meta-analyses have been conducted in almost all subdisciplines of oncology contributing to delineate the role of each treatment modality, preparing the ground for new randomised trials and, in some instances, answering unresolved therapeutic questions.

A recent example of such meta-analyses was published in 2002 in the *Journal of Urology* by Sylvester et al. [28] concerning the role of BCG (Bacillus Calmette-Guerin) instillations in patients with superficial bladder cancer. A total of 24 trials with progression information on 4,863 patients were included in the meta-analysis. Only one very small trial suggested a significant benefit from BCG instillations for this indication. The meta-analysis demonstrated a reduction of 27% in the odds of progression for patients receiving BCG. A typical representation of the results of meta-analysis in a Forest plot is shown in Fig. 2.

While meta-analyses play a very important role in the decision-making process, they should be looked

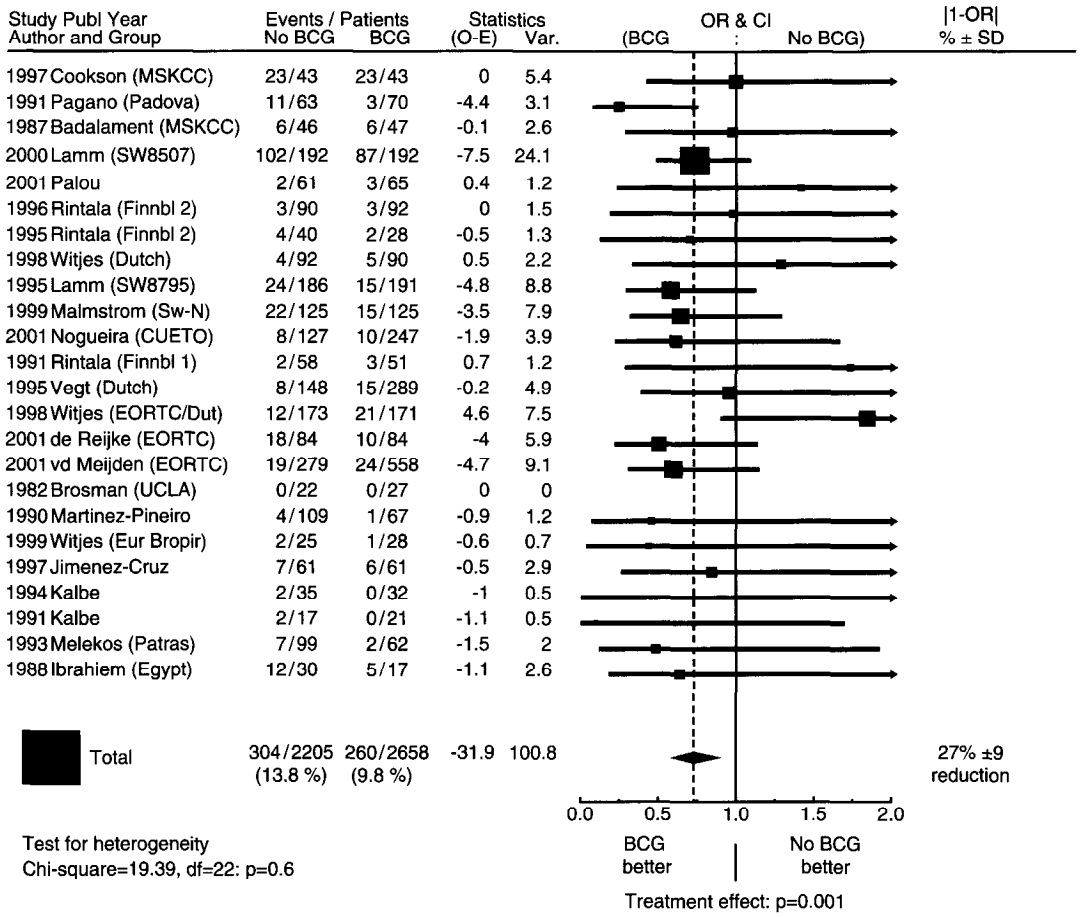

Fig. 2. Forest plot of progression by study (reproduced with permission of R. Sylvester). CI, confidence interval; E, expected; O, observed; Var., variance; df, degrees of freedom; OR, odds ratio; SD, standard deviation; BCG, Bacillus Calmette-Guerin; EORTC, European Organisation for Research and Treatment of Cancer; MSKCC, Memorial Sloan-Kettering Cancer Center.

at critically, just as most scientific methods and some meta-analyses have indeed produced results that were later invalidated in subsequent large randomised comparative trials [7]. However, to a large extent, many of these criticisms can be overcome by posing a well-formulated question, excluding improperly randomised trials, collating the individual patient data and using a basic endpoint such as survival.

Finally, as already mentioned in the introduction, one cannot talk about systematic reviews without mentioning the work of A. Cochrane and the Cochrane collaborative network, applied in all disciplines of medicine. The aim of the collaborative network is to prepare and maintain systematic reviews. The Cochrane cancer network started in 1997 and continues its development. The Cochrane database can be consulted at *www.cochrane.co.uk* [29].

Evidence-based guidelines, consensus and expert opinion

This paper would not be complete without mentioning clinical practice guidelines and other modes of scientific review used to summarise the evidence for the medical community.

In oncology in particular, due to the abundance of clinical trials and the difficulty to dissociate the good from the bad or the facts from the hypotheses, a number of official and private organisations have undertaken the development of clinical practice guidelines. The oldest and most popular in oncology is the Physician Data Query (PDQ) Database maintained by the National Cancer Institute in the USA but available to all physicians and patients around the world. It assembles all available evidence for a particular pathology and offers therapeutic guidance as well. Other national-based organisations have undertaken the same work and have published their own recommendations in national languages taking into

Table 6
Guidelines development methods

Expert opinion	Consensus	Evidence-based
Fast and cheap, unstructured	Structured, properly done is time-consuming/expensive	Systematic method, structured, expensive, and time-consuming
Possibly useful when evidence is weak	Useful mainly when uniform practice is the main goal	
Evidence considered implicitly	Evidence explicit	Explicit and generalised
"Expertise" defined by content only	Different view-points	Can be reproduced
Conflicts of interest, bias	When consensus is the main goal, it may be at the expense of evidence	Unbiased or available for scrutiny
Even experts disagree		
A biased sampling strategy for research findings	Sampling may be biased	Sampling rigorous and explicit

account standards of care recognised at the national level [1,30].

Also, under specific circumstances, the evidence is derived from consensus guidelines and expert opinion such as those organised by the National Institutes of Health (NIH) or the St. Gallen breast cancer conference. As illustrated in Table 6, the reliability of guidelines may depend heavily on the context under which they are developed [30] and, as expected, subjectivity and reliability go in opposite directions.

Conclusions

The nature of care given to patients with cancer is all too often determined by data that are more fragmentary and inconclusive than the physician would prefer. Despite rapid advances into the biology of cancer, it is also unrealistic to think there are going to be more examples of "*res ipsa loquitur*" to guide practice than there have been in the past 50 years. It is therefore incumbent on the clinical research community to pursue investigations that are based on a realistic appraisal of the findings of phase I/II trials and to support the comparison of encouraging findings in phase III comparative randomised controlled trials. The adoption of phase II results as standard, the failure to pursue phase III possibilities sooner rather than waiting for more exciting concepts to emerge, and an unwillingness to accept potential value of replicated results from phase III trials are all pitfalls that pervade the clinical research enterprise. Meta-analyses are useful but cannot replace well-conducted phase III trials, and clinical practice should be developed on the basis of the most objective evidence.

References

1 Fervers B, Bey P, Maigne D, Philip T. Standards, options and recommendations: a multidisciplinary program to promote evidence-based oncology. Bull Cancer 2001, 88: 601–604.
2 Sackett DL, Rosenberg WMC, Muir Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ Jan 1996, 312: 71–72.
3 Badenoch D, Heneghan C. Evidence-based medicine toolkit. In BMJ Books. London, BMJ Publishing Group, 2002, 50–54.
4 Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. JAMA Dec 1995, 274: 1800–1804.
5 Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ Aug 2001, 323: 334–336.
6 Imadalou K, Cvitkovic E. Phase III trials in oncology. Bull Cancer 1999, 86: 143–147.
7 Cox JD. Evidence in oncology. The Janeway lecture 2000. Cancer J, Nov-Dec 2000, 6: 351–357.
8 Djulbegovic B, Loughran TP Jr, Hornung CA, et al. The quality of medical evidence in hematology–oncology. Am J Med Feb 1999, 106: 198–205.
9 Brooks AD, Sulimanoff I. Evidence-based oncology project. Surg Oncol Clin N Am Jan 2002, 11: 3–10, vii.
10 Brooks AD, Karpoff HM, Sulimanoff I, Coit D, Brennan MF, Jacques DP. The search for level I evidence in solid-tumor oncology. Ann Surg Oncol Sept 2001, 8: 638–643.
11 Grimes DA, Schulz KF. An overview of clinical research : the lay of the land. Lancet Jan 2002, 359: 57–61. Review.
12 Tannock IF. Collaborative clinical trials: quality or quantity? Int J Radiat Oncol Biol Phys Feb 2001, 49: 339–343.
13 Hortobagyi GN. High-dose chemotherapy for primary breast cancer: facts versus anecdotes. J Clin Oncol 1999, 17 (Nov Suppl): 25–29.
14 Van Glabbeke M, Steward W, Armand JP. Non-randomised phase II trials of drug combinations: often meaningless, sometimes misleading. Are there alternative strategies? Eur J Cancer 2002, 38: 635–638.
15 Miller TP, Dahlberg S. Treatment of diffuse large-cell lymphoma: a summary of outcome for patients treated by Southwest Oncology Group. Cancer Treat Res 1993, 66: 53–63.
16 Stenning SP, Parmar MKB. Designing randomised trials: both

large and small trials are needed. Ann Oncol 2002, 13: 131–138.

17 Parmar MK, Ungerleider RS, Simon R. Assessing whether to perform a confirmatory randomised clinical trial. J Natl Cancer Inst Nov 1996, 88: 1645–1651.

18 Piccart MJ, Bertelsen K, James K, et al. Randomised Intergroup trial of cisplatin-paclitaxel versus cisplatin-cyclophosphamide in women with advanced epithelial ovarian cancer: three-year results. J Natl Cancer Inst May 2000, 92: 699–708.

19 McGuire WP, Hoskins WJ, Brady MF, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. N Engl J Med Jan 1996, 334: 1–6.

20 Muggia FM, Braly PS, Brady MF, et al. Phase III randomised study of cisplatin versus paclitaxel versus cisplatin and paclitaxel in patients with suboptimal stage III or IV ovarian cancer: a gynecologic oncology group study. J Clin Oncol Jan 2000, 18: 106–115.

21 International Collaborative Ovarian Neoplasm Group. Paclitaxel plus carboplatin versus standard chemotherapy with either single-agent carboplatin or cyclophosphamide, doxorubicin, and cisplatin in women with ovarian cancer: the ICON3 randomised trial. Lancet Aug 2002, 360: 505–515.

22 Moss C, Kaye SB. Ovarian cancer: progress and continuing controversies in management. Eur J Cancer Sep 2002, 38: 1701–1707.

23 Douillard JY, Cunningham D, Roth AD, et al. Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicenter randomised trial. Lancet March 2000; 355: 1041–1047.

24 Saltz LB, Cox JV, Blanke C, et al. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. N Engl J Med Sep 2000, 343: 905–914.

25 Mayer RJ. Moving beyond fluorouracil for colorectal cancer. N Engl J Med Sep 2000, 343: 963–964.

26 Sylvester RJ, Collette L, Duchateau L. The role of meta-analyses in assessing cancer treatments. Eur J Cancer Jul 2000, 36: 1351–1358. Review.

27 Parmar MKB, Stewart LA, Altman DG. Meta-analyses of randomised trials: when the whole is more than just the sum of the parts. Br J Cancer Aug 1996, 74: 496–501.

28 Sylvester RJ, van der Meijden APM, Lamm DL. Intravesical bacillus calmette-guerin reduces the risk of progression in patients with superficial bladder cancer: a meta-analysis of the published results of randomised clinical trials. J Urol Nov 2002, 168: 1964–1970.

29 Williams CJ. Evidence-based cancer care. Clin Oncol (R Coll Radiol) 1998, 10: 144–149.

30 Browman GP. Evidence-based paradigms and opinions in clinical management and cancer research. Semin Oncol Jun 199, 26: 9–13.