available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.ejconline.com

Position Paper

Update in methodology and conduct of cancer clinical trials

P. Therasse, E.A. Eisenhauer, M. Buyse*

EORTC Data center, Av. E. Mounier, 83 box11, 1200 Brussels, Belgium

ARTICLE INFO

Article history:
Received 31 January 2006
Received in revised form
10 February 2006
Accepted 14 February 2006
Available online 5 June 2006

Keywords:
RECIST
Interim analysis
Clinical trial directive

ABSTRACT

Many interesting changes are regularly brought into the methodology of cancer clinical trials. This position paper focuses on three topics which are felt to appear as recurrent problems which deserve more attention from the scientific community.

RECIST guidelines were published five yeas ago and have since then been largely implemented and used in clinical trials. Although the criteria were initially designed for screening phase II trials they have been used also in most phase III studies aiming at determining the efficacy of new treatments. Problems have been identified some of which require further clarifications and others deserve further research which is being undertaken. Overall RECIST is well accepted and a revised version is being considered for 2007. Interim analysis is also an important issue revealed recently through many large adjuvant or advanced trials being prematurely discontinued at the time of an interim analysis. In most instances trials were stopped because of evidence of superiority of the investigational treatment over the standard treatment. Premature discontinuation of trial poses a number of challenges addressed in this paper. Finally, the consequences of the implementation of the EU clinical trial directive are being discussed. The conclusions are without equivoque. There is much less academic research conducted in Europe, there is a lot of discrepancy and inconsistency in the implementation of the directive across member states and there is no apparent direct benefit for the patients.

1. RECIST Revisited

1.1. *Background*

Clinical cancer research takes place in an international arena: results of studies in North America can influence research and practice in Europe and vice versa. Thus, those engaged in clinical cancer research require a common "language" to describe trial outcomes. In *early drug development*, objective changes in tumour size are used to describe the impact of new agents in solid tumours. These so-called "response criteria" were first developed over 30 years ago, with those of the World Health Organization (WHO), based on summation of the products of bidimensional measurement of tumour lesions, the most commonly used for the next 2 decades.[1]

In 2000, an international working group with membership from European and North American cancer research organizations published[2] new response criteria (Response Criteria in Solid Tumours; RECIST criteria), which have subsequently been widely adopted, largely supplanting the use of the WHO criteria. The RECIST working group had in mind several goals in re-visiting response criteria; amongst them to clarify, to simplify and to provide definitions when these were missing. Thus in RECIST, definitions of measurable and non-measurable

* *Corresponding author:* Tel.: +32 2 774 16 14.
E-mail address: patrick.therasse@eortc.be (P. Therasse).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2006.02.006

disease were given, advice on methods of measurement offered, the minimum number of lesions to follow were proposed (10 maximum overall but no more than 5 per organ/site), and new definitions for standard response categories of complete response, partial response, stable disease and progressive disease were created, that took advantage of data showing that the sum of unidimensional measurements of lesions provided similar outcomes (with greater simplicity in calculation) to the sum of bidimensional lesion products.

1.2. *Application of RECIST criteria*

As was pointed out in the original paper, RECIST criteria were developed for use when objective response is the primary end point of the trial, i.e. for the most part in phase II trials of new agents, where objective response rates above certain predefined thresholds signalled biological activity of the new agent, and deemed worthy of pursuit in further clinical development. In fact, in this setting, objective response is a useful end point; agents producing objective responses in phase II trials above 10–20% are often (but not always) shown in subsequent development to lead to improvement in survival.

1.3. *Implementation issues with RECIST criteria*

Since the RECIST criteria were implemented in 2000, others have confirmed in prospective analyses the validity of substituting unidimensional for bidimensional (and even three-dimensional)-based criteria.[3–7] However, a number of questions and issues have arisen. These include: how to apply RECIST criteria in randomized phase III trials where progression-free survival is the primary end point and response?; can fewer than 10 lesions be assessed?; are RECIST criteria applicable in trials of non-cytotoxic drugs?; and what about newer imaging technologies such as PET and MRI?

1.4. *Use of RECIST in randomized trials*

As noted above, RECIST were developed for use in clinical trial setting where objective response was the primary end point. In randomized phase III trials, it is unusual that response would be the primary end point: more often it is a secondary end point. In these circumstances, RECIST criteria make provision for altering some of the usual requirements to accommodate the larger patient numbers and the logistics of applying RECIST as originally proposed to all enrolled patients. As noted in the original RECIST publication: "in phase III trials where response is not the primary end point it might not be necessary to measure as many as 10 target lesions, or to confirm response with a follow-up assessment after $\geqslant$4 weeks. Protocols should be written clearly with respect to planned response evaluation and whether confirmation is required so as to avoid post-hoc decisions affecting patient evaluability." Thus, it is possible, using the original RECIST publication as reference, to modify the criteria as noted, but the modification(s) must be described in the protocol, must apply to all study arms and should not be derived post-hoc.

Another area of difficulty in applying RECIST in phase III trials relates to the definition of 'progression'. Increasingly, progression-free survival (PFS) is being utilized as the primary end point in trials in patients with metastatic or advanced disease. The rationale for this is based on several factors: firstly, progression-free survival is a surrogate for overall survival in some settings, secondly, progression antedates death, often substantially, so the use of PFS provides an earlier indicator of whether a new treatment offers an advantage over survival, and finally, it is argued that increasing the time to progression is plausibly related to prolongation of symptom-free interval, thus justifying its use on the basis of quality of life considerations.

Whatever the reason, the requirement that all patients be assessed for progression means that there must be a standardized approach to measure this for all enrolled patients. The problem in the use of RECIST relates to the population that one might wish to include in these randomized trials: usually those with measurable or non-measurable disease at baseline. Assessing progression in patients with measurable disease is, of course, feasible since it can be based on the RECIST criteria as written. Assessing progression in those who do not have measurable disease at the time of study entry is more difficult: by definition assessing, or even defining, what level of increase in tumour burden constitutes progression, when by standard use of the term, the patients don't have quantifiable disease at baseline is problematic. Although this is seen as a problem that is RECIST-related, in fact, all response criteria suffer from the same problem: how do you assess measurable progression when there is no baseline measurement from which a minimal increase can be calculated? When such patients have new disease identified, it is easy, but what if they have simply "increasing" disease? What level of increase means they have crossed the line and should be called progressive? RECIST did include some language to address aspects of assessment of non-measurable (or non-target) disease by including the somewhat vague footnote that "Although a clear progression of "non-target" lesions only is exceptional, in such circumstances, the opinion of the treating physician should prevail and the progression status should be confirmed later on by the review panel (or study chair)." In other words, when there is a dramatic increase of non-target (or non-measurable) disease such that it is obvious that progression has occurred, this provision says it is reasonable to consider the patient progressive. Remember, however, that this footnote was written for use when the patient also had measurable disease. Furthermore, it was intended to be a rare circumstance where it would be invoked, and, finally, it was subject to external review. This is not at all the situation we find ourselves in when large numbers of patients are enrolled on randomized trials.

So what are the options? We could use PFS as a primary end point but only allow entry of patients who have measurable disease and thus can be assessed reliably using RECIST criteria. Alternatively, we could revert to the use of survival as the primary end point and allow broader patient entry, defining progression (as a secondary end point) on a trial-specific basis for those who have no baseline measurable disease.

Neither of these may be satisfactory; in fact some have called for a wider discussion of the definition of progression to be applied in randomized trials, including how and when it should be assessed (to avoid the bias that may occur when investigations are not undertaken at the same time points in

study arms). This initiative would be a welcome one since, whether one uses RECIST, WHO or one of a number of other criteria, the issue of how to assess measurable progression when only non-measurable disease is present is one which is not dependent on the response criteria used, but on the contradiction inherent in the setting.

1.5. *Minimum number of lesions*

Soon after RECIST were published, one of the most vocal objections to the use was raised: the requirement to assess a maximum of 10 lesions (or 5 per organ).[8,9] In fact, this requirement was not based on data but on what was seen as likely to be a number beyond which little further information would be gathered on assessment of total tumour burden. It seemed unreasonable to declare a patient who had 9 or 10 lesions scattered throughout their body assessable on the basis of one or two, selected on the basis of no particular criteria. However, it may well be the case that, even when 10 or more lesions are present, the same conclusions would be drawn if the target lesions identified for follow-up numbered 3 or 4 or 5-those that were largest and representative of all involved sites. What has been lacking are the data upon which it would be possible to recommend a change to the RECIST criteria in this regard. A well conducted review of real patients who had 8–10 lesions to follow is required. Using such a database the actual number of lesions above which no further information was gained in response conclusions could be shown, and would be an excellent project around which to base a modification of RECIST. A project like this is, in fact, is planned by the RECIST working group, so it can be hoped that modified criteria with a reduced maximum number lesions based on evidence from a real patient data set will be forthcoming soon.

1.6. *Use of RECIST in phase II trials of "non-cytotoxic" drugs*

It has been argued that newer non-cytotoxic agents, which will perhaps control tumour growth rather than cause objective response, may need different end points to screen for efficacy in phase II trials. RECIST criteria were modeled on older objective response criteria where partial responses in particular were identified as a "marker" of agents that might be effective in improving survival. One of the more eloquent papers written to express this point of view was published last year by Eckhardt and Ratain as an editorial entitled: "Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST."[10]

The main point of their argument was that end points and designs to detect activity of targeted agents may not be those of cytotoxic drugs. In particular they state: "Overall, we need to be more flexible in our end points and our definitions of antitumour activity, as long as the effect is distinguishable from no treatment or a placebo. This is really an issue to be addressed in phase II, as a mechanism to avoid failures in phase III. The RECIST criteria for response (and its predecessors) were designed primarily for cytotoxic agents and are not applicable to all new agents. For example, these criteria do not consider durable modest regressions or prolonged disease stability as activity, which we now know is an effect of several agents such as gefitinib, erlotinib, and bevacizumab. On the other hand, we should not rush to falsely define drugs as active on the basis of stable disease, since stable disease is a composite outcome consisting of inherent tumour growth kinetics and potential drug effect."[10]

In fact RECIST criteria, like others before it, simply described how to categorize objective observations of tumour size over time. The criteria themselves do not declare what level of response, duration of stable disease or rate of progression-free survival should "signal" that a new agent has activity worth pursuing further. Focusing the argument on the criteria used to categorize tumour size is flawed: the way we describe what happens is not the problem, it is what distribution of those categories and their frequency should signal that a new drug shows promise. For example, if we learn that non-cytotoxic drugs with single agent response rates in the 10% range do produce survival advantages in phase III studies (as is the case for erlotinib, bevacizumab), then the design of screening trials need not be based on "new end points' but rather should deploy statistical methods to detect response rates in that range (i.e. Ha should be 10% not 20% in standard design parlance). Furthermore, end points in which the composite of response + stable disease of prolonged duration are possible, as is remaining progression-free for a minimum period of time, and do not require new response criteria per se (since both of these parameters may be assessed using RECIST criteria), just new approaches to declaring what the end point is of interest, and what frequency of observation merits further investigation of the new drug.

1.7. *Imaging methods*

Positron emission tomography (PET), PET/CT, and magnetic resonance imaging (MRI) are used with increasing frequency to assess disease status and to measure its size. In Appendix V, the RECIST publication provides a commentary and guidelines on imaging methods. Ultrasound, being operator dependent, is not advised and CT scans are suggested as preferred since they are operator-independent and were the most widely-used technique in 2000 when the criteria were published. MRI is discussed as well but some issues were raised with its use as follows: "... MRI scanners vary in the images produced. Some of the factors involved include the magnet strength (high field magnets require shorter scan times, typically 2–5 minutes), the coil design and patient co-operation. Wherever possible, the same scanner should be used. For instance, the images provided by a 1.5T scanner will differ from those using a 0.5T scanner. Although, a comparison can be made, it is not ideal. Moreover many patients with advanced malignancy are in pain, so their ability to remain still for the duration of a scan sequence, in the order of 2–5 minutes is limited. Any movement during the scan time leads to motion artifacts, degradation of image quality such that the examination will probably be useless." However, if the issues raised by the foregoing are accounted for, sequential MRI images can certainly be used to follow disease size.

PET/CT is also acceptable since it provides excellent anatomical images, but intravenous contrast should be used to allow consistent and accurate tumour size measurement.

The principles behind the use of imaging modalities to assess tumour size are few, but should be adhered to: the modality used should not be operator dependent; it should allow detection of a size that is half or less of the required minimum for trial entry (so if trial requires 10 mm size, the modality should be accurate to 2–3 mm, in order to detect changes of the magnitude required for response and progression), and the baseline and follow-up images should be possible to obtain using the same technique.

The above comments relate to use of these techniques to measure anatomical tumour size and location. But what about functional measures of tumour metabolism, blood flow and other parameters now possible with PET, MRI and CT? RECIST does not address these at all. The use of new imaging modalities and the capabilities of some of these to detect functional, as well as anatomical, assessment of disease mandates that the RECIST working group, or similar group, consider how best to deploy these techniques in objective assessment of tumour behavior under treatment. While there will remain a place for objective tumour response as an end point for some time in the future, the added value (or perhaps event the replacement value) of functional changes in tumour metabolism or other measures requires integration in order to standardize the description of results of trials that incorporate these measures.[11,12]

1.8. Summary

As long as studies in patients with advanced or metastatic disease are undertaken, we will need to be able to describe the change in the size of the tumour that occurs when treatment is given using standard criteria. RECIST criteria address many issues in response assessment requiring standardization and offer practical advice which remains useful today. Issues of how to measure progression in randomized trials that enroll patients with non-measurable disease, and how best to incorporate the use of functional imaging, both require further consultation and discussion with relevant participants to find solutions to the issues they present to trial conduct and to description of trial results.

2. Simple methods for interim analysis

2.1. Background

Interim analyses are increasingly planned in cancer clinical trials in order to stop them as soon as[13] serious safety concerns emerge such that the risk / benefit ratio is unlikely to be favourable,[14] a treatment benefit is established reliably, or[15] it is clear that no treatment effect emerges and it would be futile to continue entering patients in the trial.[16] Other purposes of interim analyses may be to suggest changes in the design of the trial, such as changing patient inclusion criteria, dropping an arm, adapting a dose, or even modifying the number of patients required (also known as the "sample size" of the trial). Recent advances in statistical methodology and in trial organization make it possible to design and conduct trials with adequate provisions for such interim analyses. In most trials of long duration, interim analyses of safety and efficacy are performed at regular intervals for review by an Independent Data Monitoring Committee (IDMC). Although interim analyses are sometimes exceedingly complex, we will argue that simple measures of treatment effects can be calculated and assessed for statistical significance at any time during a clinical trial. We will also argue that IDMCs can fulfill their role adequately based on such simple statistics.

2.2. A simple estimate of treatment effect

Consider a randomized clinical trial with two treatment groups A and B having accrued roughly equal numbers of patients. Suppose that at a given time, "a" events of interest have been observed in treatment A and "b" events in treatment B. A simple formula can be used to estimate the relative risk of the event:

$$RR = a/b$$

If a is larger than b, the relative risk is larger than 1 while if a is smaller than b, the relative risk is smaller than 1. Pocock has pointed out that the statistical significance of this relative risk can be calculated with the following statistic:

$$Z = (a - b)/(a + b)^{1/2}$$

Under the null hypothesis of no difference between the treatment groups, the statistic Z is normally distributed and hence the statistical significance (or "P-value") of any observed value of Z can be looked up in a standardized normal table. If there was truly no difference between the treatment groups, the value of Z would be close to zero; it would only deviate from zero as a result of the play of chance. Table 1 shows some values for Z and the corresponding P-value, which gives the probability of observing these values of Z if there was truly no difference between the treatment groups.

Two noteworthy values of Z are 2 (with associated P-value of 0.05) and 3 (with associated P-value of 0.003).

Let us take an example. In a clinical trial that randomized patients equally between two arms denoted A and B, the number of observed events at the time of analysis is 40 in arm A and 24 in arm B. At that time Z is thus equal to

$$Z = (40 - 24)/(40 + 24)^{1/2} = 16/8 = 2.0$$

Z equal to 2.0 corresponds to a P-value of 0.05, so the difference between arms A and B is just conventionally significant. The relative risk is estimated by

$$RR = 40/24 = 1.67$$

Now suppose that at the time of analysis the number of observed events is 44 in arm A and 20 in arm B. At that time Z is equal to

$$Z = (44 - 20)/(44 + 20)^{1/2} = 24/8 = 3.0$$

Z equal to 3.0 corresponds to a P-value of 0.003, so the difference between arms A and B is highly significant. The relative risk is now estimated by

$$RR = 44/20 = 2.2,$$

indicating that the risk of the event is more than doubled with treatment A as compared with treatment B.

Table 1 – Some Z-values, corresponding P-values and interpretation

Z-value	P-value	Interpretation
1.0	0.32	Not significant (1 chance in 3 of this result being a false positive)
2.0	0.05	Conventionally significant (1 chance in 20 of this result being a false positive)
3.0	0.003	Highly significant (1 chance in 300 of this result being a false positive)
4.0	0.00006	Massively significant (1 chance in 1600 of this result being a false positive)

2.3. *Interim analyses*

Sequential analysis has been known in statistics for a long time. It was developed by Wald and other statisticians during World War II to monitor the reliability of the production of weapons and munitions. Although the use of sequential methods in medicine is much more recent, they have been used successfully in clinical trials for which the outcome of interest is observed soon after treatment, so that the analysis of the trial data can be performed almost in continuous fashion.[16] This situation is rather exceptional in cancer treatment, where some patient follow-up is typically required before an analysis of the relevant end points can be contemplated. In this case, it is more convenient to analyse the data at pre-specified time points, using so-called "group sequential designs".[15] In these designs, the number and timing of interim analyses are pre-specified in the trial protocol, and a significance level is associated with each analysis in such a way that the overall significance level of the trial is preserved (say, at a conventional level of 0.05). A simple, yet very appropriate approach proposed by Haybittle and Peto consists of using a significance level of 0.001 fpr interim analyses (i.e. an interim result is declared "statistically significant" if its P-value is less than 0.001), and to use the unadjusted level of 0.05 for the final analysis, as shown in Table 2. Although strictly speaking, the level to be used in the final analysis should be corrected to account for the interim analyses, the correction is trivial and can be ignored in first approximation.

Another commonly used approach to group sequential designs was proposed by O'Brien and Fleming and consists of using increasing significance levels, as indicated in Table 3.

Table 2 – Significance levels to use in successive analyses when the Haybittle-Peto group sequential design is used with an overall significance level of 0.05, and the number of analyses ranges from 1 (no interim analysis) to 5 (4 interim analyses)

Number of analyses	1	2	3	4	5
P-value for 1st analysis	0.05*	0.001	0.001	0.001	0.001
P-value for 2nd analysis	–	0.05*	0.001	0.001	0.001
P-value for 3rd analysis	–	–	0.05*	0.001	0.001
P-value for 4th analysis	–	–	–	0.05*	0.001
P-value for 5th analysis	–	–	–	–	0.05*

* Final analysis.

Table 3 – Significance levels to use in successive analyses when the O'Brien-Fleming group sequential design is used with an overall significance level of 0.05, and the number of analyses ranges from 1 (no interim analysis) to 5 (4 interim analyses)

Number of analyses	1	2	3	4	5
P-value for 1st analysis	0.05*	0.0054	0.0006	0.00005	0.000005
P-value for 2nd analysis	–	0.049*	0.015	0.0039	0.0013
P-value for 3rd analysis	–	–	0.047*	0.018	0.0085
P-value for 4th analysis	–	–	–	0.042*	0.023
P-value for 5th analysis	–	–	–	–	0.041*

* Final analysis.

2.4. *Independent Data Monitoring Committees*

In most trials of long duration, interim analyses of safety and efficacy are performed at regular intervals for review by an Independent Data Monitoring Committee (IDMC).[13,14] While safety analyses are carried out on a regular basis to detect untoward safety signals as soon as possible, a more conservative approach is adopted for efficacy analyses, in order not to stop a trial too early, at a stage where the treatment effects are not yet well documented. A "group sequential approach" is typically used for such efficacy analyses, with the design of O'Brien and Fleming (as shown in Table 3) being the most popular. One of the drawbacks of this design is that the analyses are assumed to be equally spaced (in terms of observed numbers of events) and their number is designed *a priori*. In reality, it is often convenient to be able to carry out an interim analysis when appropriate (for instance, after disclosure of the results of another related trial) rather than when pre-specified. Lan and DeMets have suggested use of a "spending function" that requires neither the number nor the timing of the analyses to be pre-specified. Such spending functions are gaining popularity as the most flexible and rigorous approach to early stopping of trials in case of extreme efficacy. However their sophistication imposes use of specialized software, and must be carried out by a statistical office. If, in contrast, the design of Haybittle and Peto is used, then interim analyses can simply be carried out at a level of significance of 0.001 whenever they are required (as shown in Table 2).

In practice, IDMCs have to evaluate trade-offs between the possible benefits and harms of a new treatment compared with the control group. It is tremendously useful (but surprisingly seldom done) to present IDMCs with just the number of events displayed by treatment group. In a trial of early breast cancer, for instance, the events of interest would be local and regional recurrences, new primaries in the contra-lateral breast, distant metastases, second non-breast primaries, deaths, and all side effects of the treatments under consideration. Using the simple statistics introduced above and the design of Haybittle and Peto (as shown in Table 2), the relative risk can be computed and its statistical significance assessed

for all end points of interest, rather than just for the primary end point of the trial (such as disease-free survival).

IDMCs make their recommendations to continue or stop a trial based on a host of considerations, among which interim analyses are only one (key) element. Such considerations also include the following:

- Whether all beneficial effects of the new treatment have been reliably established, especially those taking a long term to develop. For instance, in trials for patients with early breast cancer, disease-free survival is often the primary end point of interest, but time to distant metastases or overall survival are also important end points to consider. At the time of interim analyses, there may be enough recurrences to assess treatment effects on disease-free survival reliably, but too few distant metastases or deaths to assess treatment effects on other efficacy end points.
- Whether all harmful effects of the new treatment have been reliably established, especially those taking a long term to develop. For instance, in trials for patients with early breast cancer, some recently introduced treatments have been associated with increased risks of bone loss (when aromatase inhibitors are given instead of tamoxifen for women with hormone positive tumours) or cardiac function loss (when trastuzumab is added to chemotherapy for women with Her2 positive tumours). While some of these effects occur early on in the course of treatment, others may take years to develop and declaring treatment benefits too early may result in many women being put at risk of potentially dangerous side-effects in the long run.
- Whether other similar trials have mature data that could be used to confirm any results seen in a given trial. It is often difficult but nonetheless desirable for IDMCs of related trials to exchange some of their analyses on an on-going basis, for any suspicious result in one trial can be scrutinized in the others, reducing the likelihood of false-positive or false-negative findings in any single trial.
- Whether other on-going trials may be affected by the decision to stop a given trial. No trial is conducted in a vacuum, and the disclosure of early results may have profound impact on similar on-going trials. For instance, when the first interim analyses of trials of aromatase inhibitors were made public, some similar trials were closed to patient accrual, while in some other on-going trials, patients demanded to be crossed over to the aromatase inhibitor treatment, making the evaluation of long-term end points far less convincing (if not impossible) in these trials.

Finally, unwelcome trends have been seen recently in operating procedures for IDMCs. One is to develop exquisitely long and detailed IDMC charters, which is neither justified nor efficient. Another, in the same vein, is to present IDMCs with exceedingly complicated interim analyses, which have to be performed by statistical offices and cannot be checked with plain common sense. A third is to send a copy of all serious adverse event (SAEs) reports in real time to all IDMC members, which obfuscates the detection of any real safety signals by drowning them in a mass of useless detail. All of these pitfalls can be avoided by presenting IDMCs with simple statistics of the type discussed above, leaving the complications where they belong, which is in considering all consequences of stopping a trial early when large treatment benefits seem to emerge at the cost of some undesirable side-effects.

3. EU clinical trial directive: the way forward for member states implementation and academic research activities

3.1. *Introduction*

The European Union (EU) Directive 2001/20/EC[17] referred to hereafter as 'the directive', sets out compulsory rules and regulations to be followed in Europe with regards to administrative and legal requirements for the implementation of Good Clinical Practice (GCP) when conducting clinical trials. The primary aim of the directive is to protect the rights of the patients enrolled in clinical trials. It also aims to improve the competitiveness of European clinical research and to ensure a certain degree of European harmonisation in the administration process for clinical trials.

On the 4th of April 2001 the directive (which is applicable to all EU member states, and requires them to achieve its objectives whilst leaving them free to choose the means of implementation) was released for publication by the European Commission (EC) Directorate General for Enterprise and Industry, with a deadline of the 1st of May 2004 for its transposition in to the national legislation of all European Member States.

Over a year after the deadline for implementation of the directive, national implementations are still ongoing, and major difficulties in the process of the initiation of the clinical trials, have occurred. These are directly linked to the implementation of the directive into the national laws of member states, and in addition to this, emerging collateral issues such as an increased concern about insurance for clinical trials and requirements from the ethical committees have also become apparent.

3.2. *General Concerns for Academic Research*

The major concern for European academic research is the lack of recognition of the role of "non commercial trials" (which can be defined as trials not conducted for registration purposes), and the need to pragmatically define Investigational Medicinal Products (IMP) including access to such agents. An IMP can be defined as a pharmaceutical form of an active substance or placebo being tested or used as a reference in a clinical trial, including products which have already received a marketing authorization, but are used or assembled (formulated or packaged) in a way different from the authorized form, or are used for an unauthorized indication. At this point in time, it is not clear following the implementation of the directive what exactly constitutes an IMP and this clearly poses a major problem for academic research. For example it would be impossible for an academic sponsor to provide for free as stated in the directive each IMP used in a clinical trial if this were to include standard medication, for example anti-emetic drugs. IMP's are not equal and one should

differentiate agents introduced for the first time in Man versus well documented agents which clinical research aims at optimizing the use for optimal therapeutic benefit. Obviously patient integrity and safety must never be jeopardized but there is clearly a need to recognize and realize that certain academic trials should not have to follow the same process as those conducted by industry for registration purposes.

The directive raises a number of important issues for academic research which have not been adequately addressed in its original published text. In short, problematic issues which have not been addressed in the directive that include: the definition and responsibility of the sponsor, drug manufacturing for marketed drugs, fees for ethics committees and member state competent authorities, the provision of drugs by the sponsor for free, as well as onsite-monitoring and an overall increased administrative workload which slows down considerably the process of clinical trials.

3.3. *Specific issues related to the implementation of the directive for non-commercial trials*

It is evident that there is a need to identify the key issues to be addressed at the level of the directive to ensure that academic non-commercial trials remain possible in Europe[18,19] While the definition of sponsor (an individual, company, institution or organization which takes responsibility for the initiation, management and/or financing of a clinical trial) can appear clear, concerns do exist for academic non-commercial clinical trials supported by in part non-academic partners for example a pharmaceutical company where the partner would become co-liable on the project. Co-liability being defined as assuming responsibility for any error that should occur on the part of the sponsor. This was extended to a common situation in academic projects where a pharmaceutical industry partner would provide an investigational medicinal product and would therefore raise the issue of the level of involvement i.e. the co-liability of the commercial partner. The recently published "questions and answers" document which was issued by the EC in January 2005 clarifies this, in line with the plea of academic research by defining a non-commercial trial as a trial which is not part of a marketing authorization, where the researchers own the database and for which they are in full control of the design, conduct record and reporting. This 'questions and answers' document also clearly states that industry supported projects do not make the industry co-liable. This was a necessary statement to ensure that industry would continue supporting investigator driven academic studies.

Another, not yet fully clarified concern linked to sponsorship is whether multiple European sponsors with the condition of one and only one per Member State will be acceptable or whether one single European sponsor will be imposed per clinical trial.

Shelf access to an IMP without the necessity of specific labelling should be made possible when a marketed IMP is being evaluated for optimized approach in certain therapeutic areas such as oncology. The directive currently leaves this as a possibility which is left to the discretion of the Member States with the result that national interpretations differ according to the cultural background regarding clinical research.

Since the transposition of the directive into national laws, practical experience can best highlight the pitfalls and problems encountered, as well as the positive effects.[20] However, this expertise is still partial as not all Member States have their legislation in place and when in place, it is often common that all implementation aspects regulated by national decrees may not still be available. Another concern is that in some countries, the legislative documentation may not be available in part or total in English, as is the case in Poland and Italy. Language issues are also of concern for large networks such as the EORTC where the coordination of clinical trials is done from a central location in Brussels for obvious economical reasons. This may be a lesser concern to the industry sector which can either have national branches or afford the services of contract research organizations acting on their behalf.

3.4. *Regulations and competent authorities*

Following the new directive, competent authorities must comply with a maximum of 60 days for authorization of a clinical trial. In some Member States, the implementation of the directive does result in significant major changes of procedures with a direct impact on human resources and budgeting. Some member state competent authorities are not yet fully equipped to comply with the requirements and this has resulted in significant delays in obtaining the necessary approvals. This was especially apparent in the first few months following the implementation of the directive. Table 4 which is based on 8 EORTC trials activated since the directive implementation, indicates the number of clinical trials submitted in the various member states and the mean time required obtaining member state competent authorities and ethics committee's approvals.

Other concerns have appeared in Spain, Poland and Slovenia whereby the competent authorities have requested that clinical trial application forms should be completed in the national language. This can generate significant delays and an increase in resources, especially for academic non-commercial sponsors. Consequently, it is likely that some international academic trials will no longer take place anymore in these countries.

Table 4 – Time to activation of clinical trials (EORTC experience based on 8 trials activated in 11 countries)

	Number of Protocols	Average time to trial activation in months (ranges)
Belgium	(6x)	7.5 (4–12.6)
Germany	(5x)	8.4 (2.4–22.6)
The Netherlands	(5x)	4.5 (2.1–8.8)
France	(3x)	5.6 (2.6–8.5)
Italy	(3x)	10.2 (2.1–17.8)
United Kingdom	(3x)	5.3 (2.3–7.3)
Switzerland	(2x)	11 (5.3–16.7)
Poland	(1x)	9.9
Portugal	(1x)	6.2
Slovenia	(1x)	6.1
Spain	(1x)	8.4

Another aspect of concern is the cost issues generated by the application as the cost per application has risen in a few years from nothing to a few hundred Euros on average, and in some instances even to several thousand Euros such as in Germany. Waivers of these fees for non-commercial research are usually not granted and this therefore is another limiting factor. It is not yet clear if all EU member states will implement application costs or how these costs will differ throughout the EU, however this fact clearly places an additional financial burden on academic research, which may limit the number of trials that could otherwise be conducted.

3.5. *Ethics committees*

One of the positive aspects of the directive is that it imposes a central process for ethics committees in each country. This is referred to in article 7 of the directive where it states that member states shall establish a procedure for the adoption of a single opinion (one Ethics Committee approval covering all participating sites in a particular member state) regarding ethics committees. However although the directive has now been implemented in the vast majority of EU states, France is the only country which has a single opinion procedure for Ethics Committees in place which already existed prior to the directive. This does not mean that other member states do not have a single opinion system, but that the process by which they operate is not necessarily a synonym of single national ethical review. Indeed, the most frequent scenario is a system where by local ethical committees are also consulted and for which a central ethical committee would be responsible for the final statement of approval or rejection. So this could be interpreted as a lack of precision of the directive. Normally the central ethics committee is selected on the basis of the geographical location of the national coordinating investigator in the member state who has the responsibility to coordinate the submission with support from the sponsor to the central ethics committee. The submission process to all ethics committees in a given country can be sequential or simultaneous, but in other countries the time period allocated to local ethics committees for their agreement after the central approval has been obtained is as short as 5 days and therefore there is a need to submit the clinical trial in parallel to all Ethics Committees so that they can be prepared to validate or not the central position. Another unclear aspect is how data is transmitted between central and local ethics committees. Needless to say that over a year after the implementation of the majority of the national laws a large proportion of ethics committees are still not prepared to face this additional burden. Sponsors have thus taken over this challenge, resulting in additional work which in turn requires more staff.

Over the past few months, the costs for submission to ethics committees have also dramatically increased. As the role of the ethics committees is now legally binding, there is a need to organize and allow for more resources reflected on application costs. The burden on non-commercial research is not negligible as all the ethics committees may not have anticipated possible discount or waivers for non-commercial research.

Ethics committees have also been increasingly looking into the matter of insurance policies, specific insurance conditions and certificates which are issued by the sponsor's insurance company and depict the insurance conditions for a particular clinical trial. This topic was ignored by all a few years ago, but has now become a primary field of interest and a source of problems. What is astonishing is that the field of clinical trial insurance is extremely complex. It is at the corner stone of ethics, economy, laws and sciences and involves different professional opinions that do not necessarily understand each others correctly. Being trained in this field is unlikely to happen other than learning from practical experiences. Ethics committees are untrained and have poor views on the insurance system and functioning resulting in unreasonable demands that delay study activation. The problem results in the inability of ethics committees to appropriately read an insurance certificate, requesting therefore unjustified access to full policies and more documentation delaying the process of trial approval. This ultimately results in considerable differences in each member state with regard to insurance policies and results in an increase of clinical trial insurance costs as illustrated in Fig. 1.

At the level of ethical committees as with the MSCA, sponsors are also faced with a multiplicity of application forms, which may or may not resemble the approved European format and which are required in a local language exposing once more, non-commercial sponsors to additional difficulties.

3.6. *Future perspectives*

High-quality clinical academic research requires the involvement of numerous partners including academic networks, national regulatory authorities, ethical committees, national funding organizations, pharmaceutical industries, patients' advocacy groups, the EC and charities. Today, in order to fully comply with all aspects of the new requirements for the clinical trials legislation, it is essential that all actors involved at the various stages of the clinical trials collaborate on an international scale and work together towards a full integration and harmonization of procedures which is not the case presently.

In today's society academic clinical trials remain key studies to address public health issues and improve patient survival as these enable to establish state of the art treatments and have changed clinical practice while optimizing scarce financial resources. These initiatives, usually supported nationally by charities may be compromised when performed at the international level in such a stringent new regulatory environment. It is important to stress that clinical research is just as important as fundamental research for the progress

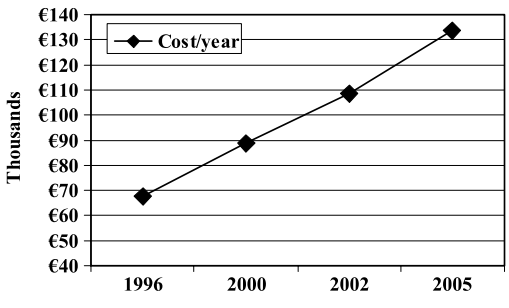

Fig. 1 – Insurance costs related to patients in EORTC trials between 1996 and 2005.

of medical science. Academic clinical trials are vital to ensure the independence of research and this ensures that clinical investigation and research remain at the service of public heath for competitiveness within the European research area.

Conflict of interest statement

None declared.

REFERENCES

1. Miller AB, Hogestraeten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;**47**:207–14.
2. Therasse P, Arbuck SG, Eisenhauer EA, et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumours. *J Natl Cancer Inst* 2000;**92**:205–16.
3. Park JO, Lee SI, Song SY, et al. Measuring response in solid tumours: comparison of RECIST and WHO. *Jpn J Clin Oncol* 2003;**33**:533–7.
4. Sohaib SA, Turner B, Hanson JA, Farquharson M, Oliver RTD, Reznek RH. CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. *Br J Radiol* 2000;**73**:1178–84.
5. Trillet-Lenoir V, Freyer G, Kaemmerlen P, et al. Assessment of tumour response to chemotherapy for metastatic colorectal cancer: accuracy of the RECIST criteria. *Br J Radiol* 2002;**75**:903–8.
6. Werner-Wasik M, Xiao Y, Pequignot E, Currant WJ, Hauck W. Assessment of lung cancer response after nonoperative therapy: tumour diameter, bidimensional product, and volume. A serial CT scan-based study. *Int J Radiat Oncol Biol Phys* 2001;**51**:56–61.
7. Therasse P, Le Cesne A, Van Glabbeke M, Verweij J, Judson Ifor the EORTC Soft Tissue and Bone Sarcoma Group. RECIST vs WHO: Prospective comparison of response criteria in an EORTC phase II clinical trial investigating ET-743 in advanced soft tissue sarcoma. *Eur J Cancer* 2005;**41**:1426–30.
8. Hillman SL, Sargent DJ, An MW, et al. Evaluation of RECIST criteria in determining the response to treatment in solid tumours: A North Central Cancer Treatment Group (NCCTG) investigation. *Proc Am Soc Clin Oncol* 2003;**22**:521. (abstr 2095).
9. Schwartz LH, Mazumdar M, Brown M, Smith A, Panicek DM. Variability in response assessment in solid tumours: effect of number of lesions chosen for measurements. *Clin Cancer Res* 2003;**9**:4318–23.
10. Ratain MJ, Eckhardt SG. Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. *J Clin Oncol* 2004;**22**(Nov 15):4442–5.
11. Kelloff GJ, Krohn KA, Larson SM, et al. The progress and promise of molecular imaging probes in oncologic drug development. *Clin Cancer Res* 2005;**22**:7967–85.
12. Kelloff GJ, Hoffman JM, Johnson B, et al. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res* 2005;**8**:2785–808.
13. DeMets DL, Furberg CD, Friedman L. *Data Monitoring in Clinical Trials. A Case Studies Approach*. Heidelberg: Springer Verlag; 2005.
14. Ellenberg SS, Fleming TR, DeMets DL. *Data Monitoring Committees in Clinical Trials. A Practical Perspective*. New York: John Wiley & Sons; 2002.
15. Jennison C, Turnbull B. *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, FL: Chapman & Hall/CRC; 1999.
16. Piantadosi S. *Clinical Trials A Methodologic Perspective*. New York: John Wiley & Son; 1997.
17. Directive 2001/20/ECof the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products. Official Journal of the European Communities 1.5.2001.
18. Lacombe D, Rea LA, Meunier F. New EU legislation may hinder academic clinical cancer research. *EJHP* 2003;**3**:14–8.
19. Baeyens AJ, Lacombe D. Regulatory issues for clinical trials at EORTC: the way forward. *Eur J Cancer* 2002;**38**:S142–6.
20. Debbaudt A, Roche L, Lacombe D, Baeyens AJ. EORTC's Regulatory Affairs Unit; Facilitating European Cancer Research. *Applied Clinical Trials*(Sept):61–6.