A new challenge in clinical research in childhood ALL: The prospective meta-analysis strategy for intergroup collaboration

M. G. Valsecchi & G. Masera

Istituto di Statistica Medica e Biomotria, Università di Milano, Milano and Clinica Pediatrica, Università di Milano, Ospedale S. Gerardo, Monza, Italy

Summary

We consider the problems arising in clinical research on childhood acute lymphoblastic leukemia (ALL). Given the therapeutic progress achieved over the last few decades, any improvement in the outcome for the majority of children with ALL is difficult to assess with the usual size trials. Furthermore, the progress in genetics and molecular biology has now led to the identification of subgroups of children, typically with rare characteristics, for whom new treatments still await evaluation. For both these aspects of clinical research, there is an increasing need for international intergroup cooperation.

After a discussion on the role of retrospective meta-analysis and randomized controlled trials in ALL research, we suggest that intergroup studies could be made more feasible, but still scientifically rigorous, by adopting a strategy of prospective meta-analysis. This strategy can be described as follows: i) different groups prospectively plan to ask the same randomized question within their protocols which may differ in other aspects, and to pool their data in order to evaluate treatment effect; ii) the management of the study can be de-centralized, by allowing each group to be responsible for conducting its own protocol.

We would like to stimulate the debate on the methodological and practical aspects of research perspectives in ALL (and in pediatric oncology).

Key words: intergroup trials, pediatric oncology, prospective meta-analysis, randomized trials

Introduction

Childhood acute lymphoblastic leukemia (ALL) is a rare disease (about 30 cases/year/million children aged 0–14 years) in which a remarkable therapeutic progress has been achieved over the last three decades. Event free survival (EFS) at 5 years from diagnosis has improved from 5% or less in the sixties to an actual 60%–70% in most cooperative groups [1], with a survival between 70% and 80% at 5 years.

For the large majority of children with ALL, who experience a relatively good prognosis, any possible improvement, based on the therapeutic resources now available, is likely to be moderate, and so can only be demonstrated with large trials. Furthermore, the progress in genetics and molecular biology has now lead to the identification of subgroups of patients whose specific biologic characteristics may have strong implications on the therapeutic approach. A relevant example is acute promyelocytic leukemia. The fact that these subgroups are typically a small part of the patient population can hinder the development of clinical research.

From a general point of view, clinical research in ALL raises the following issues:

i) how to produce new scientific evidence in a reasonable period of time?

ii) how to avoid the introduction into clinical practice of procedures whose validity is not scientifically proved? Literature is full of reports on preliminary results drawn from studies that have inadequate size and follow-up (the so called 'SNAFU syndrome': small number and follow-up);

iii) how to accelerate the spread of new important therapeutic improvements within the international scientific community?

Up to now, good quality research has been carried out with randomized controlled trials (RCTs) conducted by national or multicenter groups that coordinate a large number of centers. RCTs remain the best and most feasible research approach in countries with a large population (typically the USA), but may have limited application in countries that, in spite of high scientific and cooperative standards, can only rely on limited recruitment. This applies to various European countries, which have become increasingly aware of the need to cooperate. In 1986, the I-BFM-SG (International BFM Study Group) was created with this aim: after 10 years of fruitful scientific exchanges [2], the
conduction of a common RCT has remained a challenge.

At this stage, a discussion of the two reference sources of information on clinical intervention, i.e., meta-analysis (MA) and RCTs, as applied to the problem of childhood ALL, could clarify some issues related to future research development.

**Meta-analysis (MA)**

MA allows us to learn as much as possible from the past experiences by combining and reviewing data from well designed and closed trials. In general, MA can:

i) resolve conflicting research results by giving more precise estimates of treatment effect and by increasing the statistical power for subgroup hypothesis testing;

ii) favor the definition of a hierarchy of clinical questions worth investigating in future studies.

MA has the limitations of all retrospective evaluations: the lack of prospective planning and standardization of treatment procedures, of data collection and of analysis throughout the trials. However, it is recognized that, in areas such as oncology and cardiology, MA has made relevant contributions to the development of medical interventions [3–5]. Limitations and merits of MA have been extensively discussed in the literature [6, 7, just to mention the most recent general publications]. There are some peculiarities, in the research on ALL, which should be accounted for while discussing the role of MA. In ALL, protocols have been rapidly evolving over the last few decades, both in terms of treatment schedules, which have become increasingly complex, and of success in the outcome. Due to the complexity of treatment schedules, many similarities in drug combinations and doses may be present in similar therapeutic strategies. Randomised questions can thus be grouped by type of question and are rarely the same in different protocols. This gives rise to additional problems in the interpretation of results of meta-analysis and in drawing quantitative rather than just qualitative conclusions on treatment effects. Another drawback is that MA cannot reflect the most recent experience of the groups, as the duration of a study on ALL is usually quite long (results of a 4-year recruitment study are available after a minimum of 3–4 years from the end of recruitment). A recent MA in childhood ALL presents some of these problems and yet has contributed toward confirming the impact of duration of maintenance and treatment intensification on prognosis [8].

**Randomized controlled trial (RCT)**

RCTs are the 'gold standard' for research on clinical intervention. It is well known that they require that all participating centers:

i) adopt the same protocol for diagnosis, therapy, data collection and analysis;

ii) centralize their data in a single center entrusted with the trial management.

The practice of RCTs is well consolidated in multicenter national groups, but may be difficult to transfer to a level of international intergroup studies. Nonetheless, only intergroup studies can enable us to address significant questions on subgroups of patients by guaranteeing a rapid accrual of a large number of patients. In this respect we need to explore new perspectives in the methodology for conducting intergroup prospective studies.

**A challenge for future research**

Alternatives to the 'classical' RCT have been pursued recently in various settings [9, 10]. Basically, they adopt the idea of planning a prospective pooling (PP) of data from trials that apply the same protocol. Unlike the MA approach suggested by Chalmers [11], the PP retains the prospective nature of a RCT, as the pooling project (which data to pool and which analyses to perform) is defined without prior knowledge of the outcome. A characteristic feature of a PP is the decentralization of data management: the trials are independently run and only a set of individual data is pooled for evaluation.

According to our experience in ALL, good intergroup research could be made feasible not only by adopting the idea of decentralization but also by taking a further methodological step.

Decentralization can be advantageous because it implies that the data centers of each group, with proper funding and well established networks for management, are not dismissed. However, it is the hypothesis of adopting exactly the same protocol that implies the main resistance to intergroup research in ALL. This is mainly due to the need of each group for maintaining its own consolidated approach to some parts of treatment. This need is not only motivated on a scientific basis, but possibly also on the increased difficulties in having funding, especially if the national protocol does not make an 'original contribution' to research. In a complex and long treatment schedule, such as that used for ALL, there can be many variants but, in spite of this, many different groups share two important aspects:

i) very similar results on EFS and survival for the same type of patients;

ii) very similar views on the type of questions that need to be asked to improve the outcome of children with ALL.

In this context we think that an intergroup trial could be prospectively designed by having different groups to agree on asking the same randomized question within protocols that can be different in other aspects. We refer to this as a prospective meta-analysis (PMA) strategy, as it relies on two points that derive from the experience in large RCTs and retrospective MA [12]:
i) the idea that, if a treatment is of some benefit, then even though these benefits will probably not be the same size in each group's trial, they will, but for the play of chance, probably tend to point in the same direction. This means that it is unlikely to find unanticipated heterogeneity in results, of the type that experimental treatment is beneficial within the protocol applied by one group and is detrimental within another one (in statistical jargon, we assume that unanticipated qualitative interactions are unlikely);

ii) the results of this large trial are reliable as far as the study is strictly randomised in each group and is then analysed and interpreted unbiasedly.

These points are strengthened by the prospective nature of the PMA intergroup study. Unlike retrospective meta-analysis, here quality standards for both treatment and study management are required to be defined prior to the start of the study and to be monitored thereafter during the study, as in 'classical' RCTs. This means that, prior to beginning the study, the collaborating groups must reach a consensus on: eligibility criteria, randomised treatments, endpoints, adverse events, procedures of randomization, collection and pooling of a set of relevant data, timing and methods of analysis (and publication procedures). This methodological homogeneity is the basis on which the PMA study is expected to produce a valid and precise estimate of the relative effect of the experimental treatment as compared with the standard one. Also statistical evaluation, based on individual data, is the same as in a classical RCT, with the additional requirement that analysis of the pooled data set is stratified by group. This means that the overall estimate of treatment effect is a sort of weighted average of the results obtained in each protocol. In the case of marked heterogeneity among results of each group, a pooled effect size cannot be treated as the final word on the subject. In this case, the point of interest would be to explore the reasons why effects are diverse and to generate scientific hypothesis for future studies [13]. As required in RCTs, this intergroup study has to refer to an independent committee that is appointed to monitor safety, oversee the results and ensure that their interpretation is not data-driven.

In conclusion, the PMA maintains the methodological rigour of a classical RCT because of its prospective nature, but admits in the meantime a sort of flexibility that is typically related to meta-analysis:

i) protocols may partly differ from each other, provided that they ask the same randomized question;

ii) study conduction is decentralized.

This approach may favour the conduction of international studies for various reasons:

i) each participating group may pursue additional internal aims in its own protocol, provided that they do not jeopardise the common randomized question;

ii) it limits the burden of centralization, both economically and organizatively, but at the same time it helps data collection on outcome and prognostic factors to become more uniform in different countries;

iii) the greater responsibility given to the collaborating groups, both in planning and in conducting a PMA trial, should increase the quality of the trial and the acceptability of its results in clinical practice.

However, all these advantages have to be assessed in practice. It is also evident that this type of study, requiring an international commitment, should be undertaken under sound scientific motivations. This approach could be useful, and is likely to be essential, for studying clinical interventions on specific aspects of therapy (i.e., CNS therapy) or on rare forms of ALL (i.e., t(9;22), t(4;11), hypodiploid ALL, resistant ALL, infant ALL, extramedullary relapses). The PMA model has recently been adopted by various groups in the I-BFM-SG (AIEOP, BFM-Germany-Austria-Switzerland, EORTC, Hungary) on a therapeutic question which was judged to be relevant in the context of BFM-like protocols. The question is the effect of pulses of vincristine and dexamethasone during maintenance for intermediate risk ALL (details on the study plan will soon be published).

We consider we have achieved our aim if this paper will stimulate the debate both on methodological challenges and practical aspects of research perspectives in ALL (and in pediatric oncology).

Acknowledgements

This work was conducted within the framework of the Consiglio Nazionale delle Ricerche (CNR-ACRO, grant nos. 96.00694 and 93.02200, PF39) and with the contribution of the Fondazione Tettamanti per lo Studio delle Leucemie ed Emopatie Infantili.

References


10. The PPP Project Investigators. Design, rationale and baseline characteristics of the prospective pravastatin pooling (PPP) project. A combined analysis of three large-scale randomized trials: Long-term intervention with pravastatin in ischemic disease (LIPID), cholesterol and recurrent events (CARE), and West of Scotland coronary prevention study (WOSCOPS). Am J Cardiol 1995; 76: 899–905.

Received 10 June 1996; accepted 12 November 1996.

Correspondence to:
Maria Grazia Valsecchi, MD
Istituto di Statistica Medica e Biometria
Via Venezian, 1
20133 Milan, Italy