

Are experimental treatments for cancer in children superior to established treatments? Observational study of randomised controlled trials by the Children's Oncology Group

Ambuj Kumar, Heloisa Soares, Robert Wells, Mike Clarke, Iztok Hozo, Archie Bleyer, Gregory Reaman, Iain Chalmers, Benjamin Djulbegovic

Abstract

Objectives To assess how often new treatments for childhood cancer assessed in phase III randomised trials are superior or inferior to standard treatments and whether the pattern of successes and failures in new treatments is consistent with uncertainty being the ethical basis for enrolling patients in such trials.

Design Observational study.

Setting Phase III randomised controlled trials carried out under the aegis of the Children's Oncology Group between 1955 and 1997, regardless of whether they were published.

Main outcome measures Overall survival, event free survival, and treatment related mortality.

Results 126 trials were included, involving 152 comparisons and 36 567 patients. The odds ratio for overall survival with experimental treatments was 0.96 (99% confidence interval 0.89 to 1.03), indicating that new treatments are as likely to be inferior as they are to be superior to standard treatments. This result was not affected by publication bias, methodological quality, treatment type, disease, or comparator.

Conclusions New treatments in childhood cancer tested in randomised controlled trials are, on average, as likely to be inferior as they are to be superior to standard treatments, confirming that the uncertainty principle has been operating.

Introduction

We analysed a cohort of consecutive trials carried out by a common funder for the development of preventive and therapeutic advances in oncology, to assess the probability that an experimental treatment for cancer in children will be superior to established treatments. We assessed how often new treatments for childhood cancer assessed in phase III randomised trials are superior or inferior to standard treatments.

We also asked whether the overall pattern of treatment successes could be predicted in advance, linking the ethical principle of equipoise or the "uncertainty principle" to the pattern. This principle states that the scientific and ethical justification for enrolment of

patients into randomised controlled trials exists only if there is substantial uncertainty concerning which of the treatments is more likely to benefit patients. We previously formulated an "equipoise hypothesis,"¹—that is, if the uncertainty principle is observed, we would expect, over time, to find no significant difference between the proportion of trials that favour experimental treatments and those that favour standard treatments. We aimed to test this hypothesis.

Methods

We evaluated a consecutive series of all randomised phase III trials (126 trials involving 152 comparisons and 36 567 patients) started and completed between 1955 and 1997 under the aegis of the US National Cancer Institute sponsored Children's Oncology Group. We decided a priori to include all completed trials up to year 2000 (the last completed trial was started in 1997). These trials were published between 1960 and 2005. We analysed data from published and unpublished trials.

Collective uncertainty

We use the terms collective uncertainty and collective equipoise interchangeably. Although we could not elicit investigators' prior beliefs about the relative merits of the treatments to be compared, we can use the outcomes of the trials to assess whether they would have been right to have followed the uncertainty principle in carrying out these trials. Three possible relations exist between the trialists' uncertainties and the outcomes observed: on average new treatments are superior; on average standard treatments are superior; or on average there is no difference in outcomes between new and standard treatments. We postulated that the third possibility is most likely. This does not mean that we were seeking to show the existence of uncertainty in each

Department of Interdisciplinary Oncology, H Lee Moffitt Cancer Center and Research Institute, University of South Florida, Tampa, FL 33612, USA

Ambuj Kumar
research associate
Heloisa Soares
research associate
Benjamin Djulbegovic
professor

University of Texas MD Anderson Cancer Center
Robert Wells
professor
Archie Bleyer
professor

Children's Oncology Group
Gregory Reaman
chairman

Indiana University Northwest
Iztok Hozo
professor

UK Cochrane Centre, Oxford
Mike Clarke
director

James Lind Library, Oxford

Iain Chalmers
editor

Correspondence to: B Djulbegovic
djulbebm@moffitt.usf.edu

BMJ 2005;331:1295-8



Description of studies and additional figures are on bmj.com



This is the abridged version of an article that was posted on bmj.com on 18 November 2005: <http://bmj.com/cgi/doi/10.1136/bmj.38628.561123.7C>

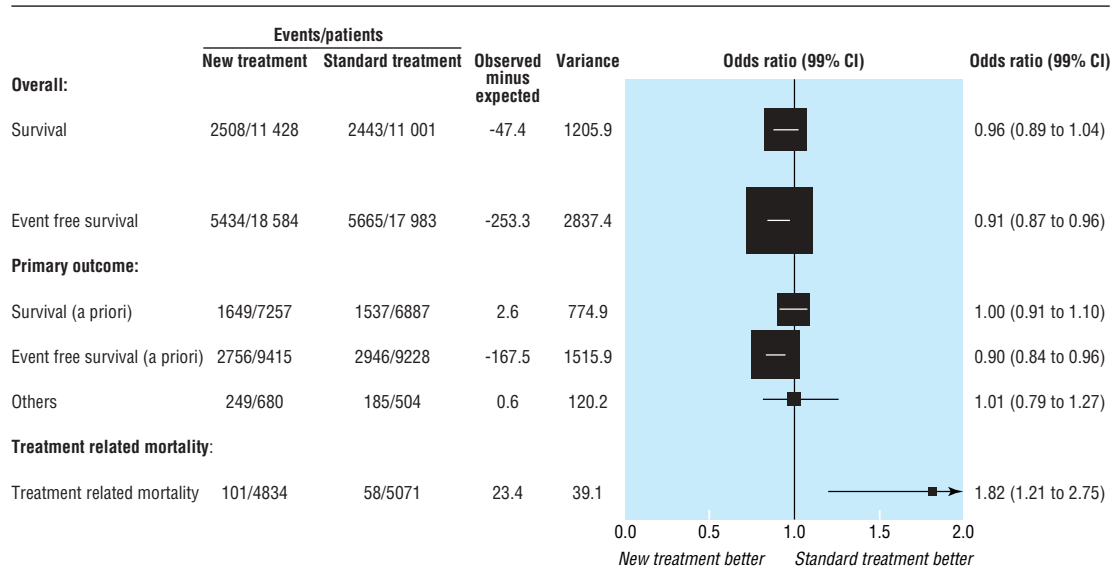


Fig 1 Summary analyses according to predefined primary outcomes. Large squares indicate trials with more information, hence narrower 99% confidence intervals. When confidence interval crosses vertical line, result is not statistically significant ($P < 0.01$)

trial. Rather, we looked for the existence of patterns over time in a large cluster of trials.^{2,3} We also examined the background to research protocols, which typically indicated that the researchers were not certain about the effect of the treatment in the trials.

Factors that may affect outcomes

Publication rate

We used the National Cancer Institute definition of completed studies to determine the publication rate: a study is considered to have been completed if it has been closed to accrual, all patients completed therapy, and the study met its primary objectives. We excluded studies that were started but were closed early owing to poor patient accrual or had not yet completed follow-up and trials that compared two new treatments ($n = 1$). Trials that were closed early because the results clearly favoured one treatment over another were included in our analysis ($n = 2$). If a study had more than one publication we extracted data from the most recent available report.

Quality assessment

We extracted data on the methodological domains relevant to minimising bias and random error in the conduct and analysis of the trials. We used both the research protocols and the final publication for each trial to assess quality. Interobserver agreements for quality appraisal and for assessment of treatment success were high ($\kappa = 0.90$ to 0.97).

Classification of comparator

The results of a trial may be affected by an inappropriate comparator. We classified comparators in the trials as either active or placebo or no active treatment, and we analysed these trials separately. Violation of the uncertainty principle relates to the choice of an inferior comparator intervention.

Distribution of outcomes between new and standard treatments

Methods used to assess the superiority of experimental or standard treatments and the distribution of outcomes are reported elsewhere.³ We assessed the proportion of trials that achieved a statistically significant difference according to the primary outcome, and the preference between new and standard treatments as judged by the original investigators.

We used meta-analysis to derive overall estimates of the likelihood and size of any average differences between the effects of new and standard treatments. The unit of analysis was each randomised comparison within a trial. In trials with more than one comparison, we biased our analysis against standard treatments, providing the best case scenario for concluding that new treatments are better (see bmj.com).

We pooled data on overall survival and event free survival. We also pooled data on treatment related mortality. Summary effects were expressed as hazard ratios or Peto odds ratios with 99% confidence intervals (see bmj.com).

We evaluated the pattern of new successes over time. If a new success influences the outcome of another success then one would expect significant correlation between experimental treatments at time t and preceding times. If, however, the uncertainty principle is at work, then testing in each trial should be independent of previous testing. Such a series of treatment successes would conform to a white noise pattern, with no significant autocorrelation in a time series analysis.

Results

We included 126 randomised trials in our analyses. After exclusions, the final analysis included 152 comparisons from 107 trials. We did not detect any differences between published and unpublished trials (see bmj.com).

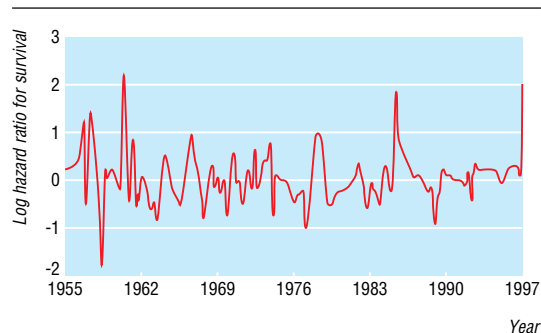


Fig 2 Time series analysis of treatment effect (log hazard ratio) of studies carried out by the Children's Oncology Group. White noise pattern indicates no significant autocorrelation between studies carried out at various time intervals. Log hazard <0 indicates superiority of new treatments and >0 a survival advantage for standard treatments

The documents we reviewed clearly indicated that, at the start of their studies, the Children's Oncology Group investigators were hoping that new treatments would be better than standard treatments. In their reports of completed studies, they stated preferences for new treatments in 47% of the comparisons and for standard treatments in 53% ($P=0.4$). The investigators judged 26 (17%) interventions as breakthrough. (We defined breakthrough interventions as those the investigators judged as highly preferred so they should become standard of care, or that had an effect size so large that their log hazard ratio for survival or event free survival was -1 or less.) Some new treatments were, however, worse than existing treatments. Only 29% of the trials achieved statistically significant differences in the investigators' a priori primary outcomes (see bmj.com).

The Peto odds ratio for the principal outcome (overall survival) is 0.96 (99% confidence interval 0.89 to 1.04; fig 1) indicating that new treatments were as likely to be inferior as they were to be superior to standard treatments (see bmj.com for forest plot showing results for individual trials). New treatments were slightly favoured for event free survival (0.91, 0.87 to 0.96), but this advantage was offset by increased treatment related mortality associated with new treatments (1.82, 1.21 to 2.75). This balance of benefits and harms is the likely reason why overall survival, on average, was similar for new and standard treatments (fig 1). We found no evidence of any autocorrelation over time between success in one trial and success in another (fig 2), suggesting that each trial represents an independent experiment in a given time with the aim of addressing the uncertainty that existed when the trial was designed.

We did not detect any association between methodological quality or choice of comparator and the outcomes of trials. This is expected because of the high quality of these trials (see bmj.com). Neither did we detect any heterogeneity within our analyses of overall results by type of cancer studied.

Discussion

New experimental treatments for childhood cancer assessed in phase III randomised trials are as likely to be inferior as they are to be superior to standard treatments. Our result also indicated unpredictability of

individual trial results. We believe that the pattern of successes and failures of new treatments is consistent with uncertainty being the basis for inviting patients to participate in such trials.

We further believe that this pattern of therapeutic success is a key reason behind major accomplishments in efforts to treat childhood cancer. The success has not come from a series of continuous, steady improvements, but from empirical testing by investigators who acknowledged their uncertainty and chose to randomise between treatments, the relative effect of which they could not predict.¹⁻³

More than 20 years ago Mosteller estimated that the public, sponsors of research, and investigators can expect that innovations will be successful about 50% of the time, which he called a "good investment."⁴ Over the past 25 years, studies of the track record of new treatments in other spheres, and outcomes in cancer in adults, have also found a similar pattern of treatment successes. Thus, a pattern seems to be emerging: when the analyses are based on complete cohorts of published and unpublished trials, identified at inception, using meta-analysis to obtain estimates that take account of trial size and time to event data, half the time new treatments are either not different or superior to standard treatments or are not different or inferior to standard treatments.

All published studies showing that innovative treatments are, on average, equally successful to standard treatments were based on publicly sponsored trials. Industry sponsored trials are associated with increased likelihood of outcomes favouring sponsors, most likely due to selective reporting of favourable outcomes and violation of the uncertainty principle in the design of the trials.⁵

The pattern we describe holds only for the overall distribution of outcomes comparing new treatments with standard treatments. This fact is the key to preserving the clinical trial system, and the willingness

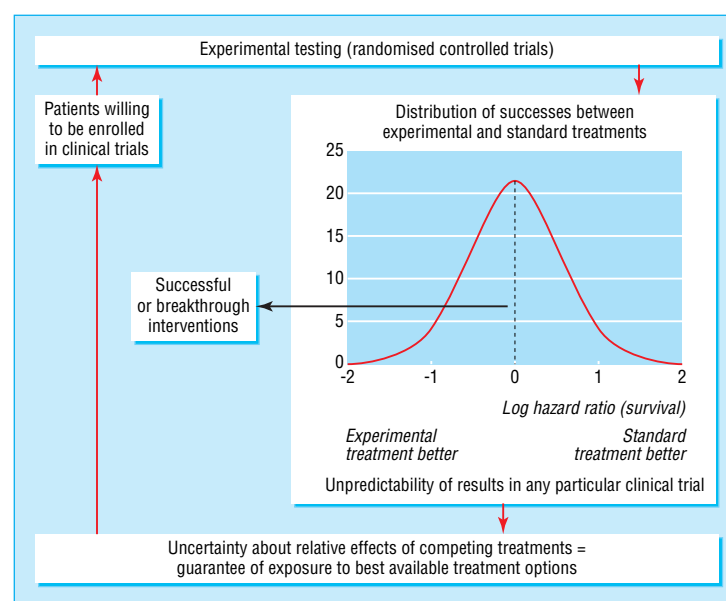


Fig 3 Proposed model of clinical discoveries showing how ethical principle converges to become scientific principle driving treatment progress. Graph insert shows actual distribution of treatment successes in trials carried out by the Children's Oncology Group (see bmj.com for details)

What is already known on this topic

Potential participants in clinical trials should be given relevant details

In paediatric oncology this should include the track record of new treatments studied in earlier trials

No analyses exist of the average track record of new treatments for childhood cancer

What this study adds

Experimental treatments for cancer in children are as likely to be inferior as they are to be superior to standard treatments

The value of new experimental treatments cannot be confidently predicted in advance

Uncertainty has provided the ethical foundation for randomised trials and has been the driver for the substantial advances in several childhood cancers

of patients to participate in clinical trials, thereby contributing to further therapeutic advances. Provided the uncertainty principle applies, there is no a priori reason to be cautious about clinical trials in general, since new treatments tend to be, on average, neither better nor worse than standard therapies. If treatment success could consistently be predicted, patients would be expected to request those successful treatments, making enrolment into clinical trials and randomisation impossible. Particular treatments may prove to be better or worse than standard treatments but this will only be known after completion of the trial. Although data from model experimental systems and phase I and II trials of novel treatments may seem promising, they do not predict sufficiently well the outcomes of the phase III trials that will inform practice.

We believe that our results are generalisable to the pattern of treatment success seen in childhood cancer, since 94% of children diagnosed with cancer in the United States are seen at institutions that are members of the Children's Oncology Group, and more than 60-65% of these patients are enrolled in the group's clinical trials.⁶

A possible limitation of our study is that we could not elicit researchers' prior beliefs about the relative merits of the treatments to be compared. However, the fact that we predicted a pattern of treatment success in advance before we collected data on treatment success⁷ provides strong corroborative evidence for our hypothesis.

Our findings, and those from similar studies, should underpin the continuing need to resolve uncertainty through the randomised comparison of new and standard treatments. The use of the uncertainty principle provides the ethical foundation for randomised trials, and has been the driver for the substantial therapeutic advances in developing effective treatments for several childhood cancers. The scientific community and the public should be made more aware of how this mechanism underlies advances in clinical medicine (fig 3).

We thank the Children's Oncology Group for their extensive cooperation, Craig Beam (Biostatistical Core, H Lee Moffitt Cancer Centre and Research Institute) for carrying out the time series analysis; William Dalton (H Lee Moffitt Cancer Centre and Research Institute), Steven Joffe (Dana-Farber Cancer Institute, Harvard Medical School), and Howard Mann (University of Utah,

Salt Lake City) for their insightful comments and helpful feedback to earlier version of this paper; Vineet Gupta (H Lee Moffitt Cancer Centre and Research Institute) for comments regarding the latest version of the manuscript; and the patients, healthcare providers, and others who took part in the clinical trials of the Children's Oncology Group and the legacy paediatric cooperative groups

Contributors: See bmj.com.

Funding: Research programme on research integrity, Office of Research Integrity and National Institute of Health (grant No 1R01NS044417-01, NIH/ORI).

Competing interests: Although some authors work within the Cochrane Collaboration and Children's Oncology Group, the views expressed here are their own and not necessarily those of the Cochrane Collaboration or Children's Oncology Group.

Ethical approval: This study was approved by the University of South Florida institutional review board.

- 1 Djulbegovic B. Acknowledgment of uncertainty: a fundamental means to ensure scientific and ethical validity in clinical research. *Curr Oncol Rep* 2001;3:389-95.
- 2 Djulbegovic B, Lacey M, Cantor A, Fields KK, Bennett CL, Adams JR, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635-8.
- 3 Soares HP, Kumar A, Daniels S, Swann S, Cantor A, Hozo I, et al. Evaluation of new treatments in radiation oncology: are they better than standard treatments? *JAMA* 2005;293:970-8.
- 4 Mosteller F. Innovation and evaluation. *Science* 1981;211:881-6.
- 5 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
- 6 Bleyer WA, Tejeda H, Murphy SB, Robison LL, Ross JA, Pollock BH, et al. National cancer clinical trials: children have equal access; adolescents do not. *J Adolesc Health* 1997;21:366-73.
- 7 Lipton P. Testing hypotheses: prediction and prejudice. *Science* 2005;307:219-21.

(Accepted 16 September 2005)

doi 10.1136/bmj.38628.561123.7C

Endpiece

Common doctors

When I say "common doctors" I mean those whom we can also call "textbook" doctors, since indeed they have their whole medicine not in their heart and mind but laid aside in books and manuscripts; and they do not know how to apply their hands to the curing of a disease unless, having first found its name, they look for the remedies under that appropriate to it in their books. Whereas true doctors, inasmuch as they prosecute their art by methods having but scant regard for names, investigate the substance and causes of diseases by division and resolution in order to elicit therefrom therapeutic indications and frame intentions by which they have recourse to books; they do not, as do the former, allow themselves to be led like blind men by names to them as judges, but putting what they read to the refinement of reason they do not obey writings as though they were masters, but rather drag the writings themselves like slaves toward the forwarding of their own purpose: the result is that this exact knowledge of names is necessary for both schools; for the former because they know no other way of healing; for the latter to enable them to enter into discussion with others and to dispute with them the method of effecting a cure.

Giovanni Manardi, *Concerning the names of disease afflicting the outward parts* (1535), quoted in Bullough VL. *Universities, medicine and science in the medieval west*. Aldershot: Ashgate Publishing, 2004:97

Peter Morrell, honorary research associate, history of medicine, Staffordshire University