

Satisfaction of the uncertainty principle in cancer clinical trials: retrospective cohort analysis

Steven Joffe, David P Harrington, Stephen L George, Ezekiel J Emanuel, Lindsay A Budzinski, Jane C Weeks

Abstract

Objective To assess whether publicly funded adult cancer trials satisfy the uncertainty principle, which states that physicians should enrol a patient in a trial only if they are substantially uncertain which of the treatments in the trial is most appropriate for the patient. This principle is violated if trials systematically favour either the experimental or the standard treatment.

Design Retrospective cohort study of completed cancer trials, with randomisation as the unit of analysis.

Setting Two cooperative research groups in the United States.

Studies included 93 phase III randomised trials (103 randomisations) that completed recruitment of patients between 1981 and 1995.

Main outcome measures Whether the randomisation favoured the experimental treatment, the standard treatment, or neither treatment; effect size (outcome of the experimental treatment compared with outcome of the standard treatment) for each randomisation.

Results Three randomisations (3%) favoured the standard treatment, 70 (68%) found no significant difference between treatments, and 30 (29%) favoured the experimental treatment. The average effect size was 1.20 (95% confidence interval 1.13 to 1.28), reflecting a slight advantage for the experimental treatment.

Conclusions In cooperative group trials in adults with cancer, there is a measurable average improvement in disease control associated with assignment to the experimental rather than the standard arm. However, the heterogeneity of outcomes and the small magnitude of the advantage suggest that, as a group, these trials satisfy the uncertainty principle.

Introduction

Clinicians and patients generally have preconceptions about the relative merits of study treatments.¹ It is therefore important to consider under what conditions of prior knowledge and belief randomised trials may ethically proceed. Many argue that “equipoise” must exist. This means that there is no “consensus within the expert clinical community about the comparative merits of the alternatives to be tested.”² Others invoke the “uncertainty principle”³ (see also bmj.com). Most authors agree that when the better treatment can be identified with reasonable confidence, it is both unethical and scientifically unnecessary to conduct the trial.⁴

Concerns about violations of the uncertainty principle, which have adverse practical and ethical consequences, are widespread.⁴⁻⁹ For a physician or patient considering a trial, deciding whether the uncer-

tainty principle is satisfied requires a set of expectations about how the new treatment compares with the standard treatment. These expectations are generally cited as grounds for determining whether offering a trial is ethical, though they may reflect inaccurate or even biased predictions about the trial's outcome. Although it is difficult to predict the outcomes of specific trials, such judgments need not be entirely subjective. Chalmers takes a bayesian perspective and considers trials as elements of a coherent system rather than as isolated events. He suggests that outcomes of completed trials can inform estimates of the “prior probability of a proposed new treatment being superior to an established treatment.”¹⁰ In the absence of bias, the expectations of physicians and patients should bear some relation to estimates of prior probabilities derived in this way.⁴ Furthermore, the finding that new treatments prove superior to standard treatments most of the time would suggest a systematic violation of the uncertainty principle. Several authors have conducted empirical studies under this premise.^{8 11 12}

We evaluated whether publicly funded US cancer trials satisfy the uncertainty principle.

Methods

Study sample—We identified phase III randomised trials coordinated by the Eastern Cooperative Oncology Group (ECOG) or Cancer and Leukemia Group B (CALGB). These two groups were among the larger multi-cancer groups supported by the US National Cancer Institute that were active during the study period. We selected studies that finished recruiting participants between 1981 and 1995. We limited the cohort to trials testing whether a new anticancer therapy was more effective than a control. We excluded studies closing early due to poor recruitment as uninformative for our purposes but included studies closing early for other reasons (for example, interim results).

Data collection—We used original protocols to assess study design. For study results, we used statisticians' reports, published articles, or meeting abstracts. We included 103 randomisations from 93 trials in our study cohort. In total, 34 943 patients were included in study analyses.

Analysis—We aimed to describe how experimental treatments in phase III cooperative group cancer trials fare when compared with standard treatments. We addressed this in two ways. In the main analysis, we calculated the proportions of randomisations that favoured the experimental treatment, the standard

Department of
Pediatric Oncology,
Dana-Farber
Cancer Institute,
Boston, MA 02115,
USA

Steven Joffe
assistant professor

Department of
Biostatistical
Science,
Dana-Farber
Cancer Institute
David P Harrington
professor

Cancer and
Leukemia Group B
Statistical Center,
Durham, NC 27710
Stephen L George
professor

Department of
Clinical Bioethics,
Warren G
Magnuson Clinical
Center, National
Institutes of Health,
Bethesda, MD
20892-1156

Ezekiel J Emanuel
director

Lesley College,
Cambridge, MA
02138

Lindsay A
Budzinski
graduate student

Department of
Medical Oncology,
Dana-Farber
Cancer Institute
Jane C Weeks
associate professor

Correspondence to:
S Joffe
steven_joffe@dfci.harvard.edu

BMJ 2004;328:1463-6



This is the abridged version of an article that was posted on bmj.com on 26 May 2004: <http://bmj.com/cgi/doi/10.1136/bmj.38118.685289.55>

Table 1 Results of comparisons between experimental and control treatments. Figures are number (percentage)

	Control better	No significant difference	Experimental better	Geometric mean effect size (95% CI)
All comparisons (n=103)	3 (2.9)	70 (68.0)	30 (29.1)	1.20* (1.13 to 1.28)
Response rate comparisons (n=42)†	1 (2.4)	33 (78.6)	8 (19.1)‡	1.11§ (1.01 to 1.22)
Survival comparisons (n=41)†	2 (4.9)	33 (80.5)	6 (14.6)‡	1.11§¶ (1.03 to 1.19)
Other time to event comparisons (n=59)†	1 (1.7)	35 (59.3)	23 (39.0)‡	1.28§** (1.16 to 1.41)

*Effect sizes could be calculated for 101 randomisations.

†Some randomisations reported two or more primary end points. Consequently, numbers for response rate, survival, and other time to event comparisons total more than 103.

‡P=0.03 for comparison of study outcome by end point (Fisher's exact test).

§P=0.01 for comparison of effect sizes by end point (one way analysis of variance).

¶Effect size could be calculated for 40 randomisations evaluating survival.

**Effect size could be calculated for 57 randomisations evaluating other time to event end points.

treatment, or neither treatment, based on significance as specified in the protocol. In a supporting analysis, we calculated an effect size (for example, ratio of median survivals or response rates) for each two way comparison. Then we estimated the average effect size for the study cohort. Finally, we conducted an exploratory logistic regression analysis to identify attributes of studies that favoured experimental treatment. We evaluated five independent variables: end point, date closed to recruitment, disease setting, placebo/observation control, and average sample size per arm. When a study reported multiple two way comparisons, we considered each as a separate data point but used clustering to account for non-independence within randomisations.

Results

Sample description

The most studied cancers were breast, small cell lung, acute myeloid leukaemia, colorectal, and Hodgkin's and non-Hodgkin's lymphomas (see also bmj.com). Recruitment required a median of 4.3 years, or 20% longer than anticipated. Seven trials closed early on the basis of interim data.

About a quarter of the trials involved three or more arms. Seventeen involved placebo or observation controls, although participants received active treatment before randomisation in 15 of these studies. The median sample size was 299 patients per randomisa-

tion, or 125 per arm. The median ratio of actual to planned recruitment was 1.0.

Study outcomes

Thirty randomisations (29%) favoured experimental treatment, three (3%) favoured standard treatment, and 70 (68%) favoured neither (table 1). The average effect size was 1.20 (95% confidence interval 1.13 to 1.28), reflecting an advantage for experimental treatment. Comparisons involving time to event end points apart from survival favoured experimental treatment more often than those involving survival or response end points, whether they were analysed as categorical outcomes or as effect sizes.

Published manuscripts were available for 86/103 randomisations (or 77/93 trials). Among published randomisations, 27 (31%) favoured experimental treatment. Among unpublished randomisations, three (18%) favoured experimental treatment (P=0.38, Fisher's exact test).

In the logistic regression analysis, the likelihood of favouring experimental treatment varied by end point, date of study closure, and disease setting, but not by type of control or sample size (table 2). Comparisons that used time to event end points apart from survival, closed in 1986-90, and involved locoregional solid tumours most often favoured experimental treatment.

Discussion

We found that, on average, experimental treatment resulted in slightly better disease control than standard therapy did. Experimental therapy rarely proved less effective than the contemporary standard.

Despite this apparent imbalance, the heterogeneity of outcomes and small average effect sizes suggest that overall these trials satisfied the uncertainty principle. However, some might disagree with this interpretation, saying that the observed difference violates the uncertainty constraint.¹³ Others might argue that given the urgent need for advances against most cancers in adults, the proportion of successful experimental treatments is disappointingly low.

What did other studies find?

Few studies, all of which have looked at published articles or abstracts, have asked related questions in oncology. Chlebowski and Lillington found that 16% of trials comparing adjuvant therapies for localised breast cancer, but only 2% of trials for advanced breast cancer, favoured experimental treatment.¹⁴ We similarly found

Table 2 Factors associated with significant difference favouring experimental treatment*†

Independent variable‡	Adjusted odds ratio (95% CI)	P value
End point:		
Overall survival	Reference	
Other time to event	4.7 (1.8 to 12.7)	0.008
Response rate	2.8 (0.8 to 9.5)	
Date protocol closed to recruitment:		
1981-5	Reference	
1986-90	4.4 (1.6 to 12.1)	0.009
1991-5	1.7 (0.5 to 5.5)	
Type of malignancy:		
Advanced solid tumour	Reference	
Haematological	1.9 (0.7 to 5.0)	0.015
Locoregional solid tumour	5.8 (1.8 to 19.0)	

*Results of logistic regression model evaluating attributes of randomisations in which compared with control arm, experimental arm met statistical criteria for superiority defined in protocol.

†Some randomisations report more than one two way comparison of experimental to control arm. Model therefore uses clustering by randomisation to adjust for lack of independence within randomisations.

‡Use of placebo/observation control and number of patients per arm were not associated with significant difference in favour of experimental arm.

that trials for advanced solid tumours were less likely than other trials to favour experimental treatment. Machin et al observed that eight of 29 (28%) UK Medical Research Council trials for solid tumours favoured experimental treatment.¹² Finally, Djulbegovic and colleagues observed that 56% of multiple myeloma trials published between 1996-8 favoured experimental treatment.⁸ This advantage was most apparent among placebo controlled trials and trials funded by industry. There were no differences by type of control in our study, though few trials lacked active controls. The higher proportion of studies with positive results in the analysis of Djulbegovic et al may reflect the prevalence of placebo controlled trials and trial funded by industry, the use of qualitative conclusions rather than hypothesis tests to define outcomes,¹⁵ the inclusion of one disease, or publication bias.

Analytic and interpretive limitations

Our analysis has several limitations. The trials we studied involved diverse patient populations, interventions, and study designs. Providing a single estimate of prior probabilities is therefore oversimplified. Also, the prior evidence favouring the experimental treatment—possibly the strongest predictor of the trial's outcome—probably varied among trials. Unfortunately, we could not quantify the weight of evidence favouring the experimental treatment that was available at the start of each trial. Nor could we formally incorporate considerations of morbidity into the analysis. Had we been able to do so, the risk-benefit ratio of the two arms would probably have been even more closely balanced as the toxicity of experimental treatment is often greater than that of standard treatment. Application of our results to current trials requires caution as temporal trends might alter the underlying probability distributions. Our data do not address the ethically important question of what physicians and others actually believed when these trials began. Studies that prospectively assessed physicians' prior beliefs and then correlated those beliefs with trial outcomes would be worthwhile. Finally, problems related to sample size could partly explain the high proportion of trials that found no significant difference between treatments.

We do not suggest that the proportions or effect sizes reported here can be used in isolation to estimate prior probabilities for individual trials. Other information about the study intervention, including biological plausibility and results of preliminary research, can indicate whether and how expectations for particular trials should differ from population norms. Nevertheless, such average probabilities can serve both as starting points for determining expectations about individual trials and to gauge the degree of uncertainty inherent in the system of trials as a whole.

Prior probabilities in other contexts, including paediatric oncology, research sponsored by industry, or non-cancer trials, probably differ from those we observed. Factors such as the inherent responsiveness of the conditions under study, sponsors' financial incentives, regulatory mandates (for example, about placebo controls), and customs regarding confirmatory trials undoubtedly affect the outcomes of trials. Such variation suggests that the uncertainty principle has ethical implications for trial systems as well as for individual studies and therefore should influence deci-

What is already known on this topic

The uncertainty principle is often cited as the ethical basis for randomised trials

In oncology, studies that evaluate whether this principle is satisfied give conflicting results and are subject to publication bias

What this study adds

In a cohort of publicly funded adult cancer trials that included unpublished studies, fewer than one third favoured the experimental treatment

On average, experimental treatments were associated with about a 20% improvement in disease control

The heterogeneity of results and small average effects suggest that as a group these trials satisfy the uncertainty principle

sion models about which therapies to advance to randomised trials. Determining the optimum distribution of prior probabilities across a system of trials is complex, involving the need to offer current patients the most effective therapies available, the obligation to avoid harming some participants by giving them experimental therapy that proves less effective than existing standards, and the mandate for efficient clinical progress. Ultimately, defining the limits of acceptable uncertainty will require that we address substantive questions about the relationships between individual patients and communities as mediated through the institution of clinical trials.

Conclusions

To summarise, we observed measurable average improvements in disease control associated with receipt of experimental rather than standard therapy in adult cooperative group cancer trials. Nevertheless, the small magnitude of the advantage and the heterogeneity of results suggest that as a group these trials satisfy the uncertainty principle. Our results may encourage recognition that prior beliefs about the relative merits of the treatments being compared are tenuous at best. By highlighting this uncertainty, we hope patients and physicians will be more willing to help to advance cancer therapy through participation in controlled trials.

We thank the Cancer and Leukemia Group B and the Eastern Cooperative Oncology Group for allowing access to primary study documents.

Contributors: See bmj.com

Funding: Greenwall Foundation and the US National Cancer Institute (CA96872). The Eastern Cooperative Oncology Group (CA23318) and the Cancer and Leukemia Group B (CA33601) receive funding from the National Cancer Institute.

Competing interests: DPH is the former group statistician of the Eastern Cooperative Oncology Group. SLG is the current group statistician and director of the statistical centre for the Cancer and Leukemia Group B.

Ethical approval: Not required.

1 Lilford RJ. Ethics of clinical trials from a bayesian and decision analytic perspective: whose equipoise is it anyway? *BMJ* 2003;326:980-1.

- 2 Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987;317:141-5.
- 3 Peto R, Baigent C. Trials: the next 50 years. Large scale randomised evidence of moderate benefits. *BMJ* 1998;317:1170-1.
- 4 Edwards SJ, Lilford RJ, Brauholtz DA, Jackson JC, Hewison J, Thornton J. Ethical issues in the design and conduct of randomised controlled trials. *Health Technol Assess* 1998;2:vi,1-132.
- 5 Hellman S, Hellman DS. Of mice but not men. Problems of the randomized clinical trial. *N Engl J Med* 1991;324:1585-9.
- 6 Fried C. *Medical experimentation: personal integrity and social policy*. New York: American Elsevier, 1974.
- 7 Marquis D. Leaving therapy to chance. *Hastings Cent Rep* 1983;13:40-7.
- 8 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.
- 9 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;327:248-54.
- 10 Chalmers I. What is the prior probability of a proposed new treatment being superior to established treatments? *BMJ* 1997;314:74-5.
- 11 Gilbert JP, McPeck B, Mosteller F. Statistics and ethics in surgery and anesthesia. *Science* 1977;198:684-9.
- 12 Machin D, Stenning SP, Parmar MK, Fayers PM, Girling DJ, Stephens RJ, et al. Thirty years of Medical Research Council randomized trials in solid tumours. *Clin Oncol* 1997;9:100-14.
- 13 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
- 14 Lee SJ, Zelen M. Clinical trials and sample size considerations: another perspective. *Stat Sci* 2000;15:95-110.
- 15 Chlebowski RT, Lillington LM. A decade of breast cancer clinical investigation: results as reported in the program/proceedings of the American Society of Clinical Oncology. *J Clin Oncol* 1994;12:1789-95.
(Accepted 2 April 2004)

doi 10.1136/bmj.38118.685289.55

Hospital admission for acute pancreatitis in an English population, 1963-98: database study of incidence and mortality

Michael J Goldacre, Stephen E Roberts

Unit of Health-Care
Epidemiology,
Department of
Public Health,
University of
Oxford, Old Road,
Oxford OX3 7LF
Michael J Goldacre
professor of public
health

Stephen E Roberts
university research
lecturer

Correspondence to:
M Goldacre
michael.goldacre@
dphpc.ox.ac.uk

BMJ 2004;328:1466-9

Abstract

Objectives To investigate trends in the incidence of acute pancreatitis resulting in admission to hospital, and mortality after admission, from 1963 to 1998.

Design Analysis of hospital inpatient statistics for acute pancreatitis, linked to data from death certificates.

Setting Southern England.

Subjects 5312 people admitted to hospital with acute pancreatitis.

Main outcome measures Incidence rates for admission to hospital, case fatality rates at 0-29 and 30-364 days after admission, and standardised mortality ratios at monthly intervals up to one year after admission.

Results The incidence of acute pancreatitis with admission to hospital increased from 1963-98: age standardised incidence rates were 4.9 per 100 000 population in 1963-74, 7.7 in 1975-86, and 9.8 in 1987-98. Age standardised case fatality rates within 30 days of admission were 14.2% in 1963-74, 7.6% in 1975-86, and 6.7% in 1987-98. From 1975-98, standardised mortality ratios at 30 days were 30 in men and 31 in women (compared with the general population of equivalent age in the same period = 1), and they remained significantly increased until month 5 for men and month 6 for women.

Conclusions Incidence rates for acute pancreatitis with admission to hospital rose in both men and women from 1963 to 1998, particularly among younger age groups. This probably reflects, at least in part, an increase in alcoholic pancreatitis. Mortality after admission has not declined since the 1970s. This presumably reflects the fact that no major innovations in the treatment of acute pancreatitis have been introduced. Pancreatitis remains a disease with a poor prognosis during the acute phase.

Introduction

Acute pancreatitis has become increasingly common in Western countries in recent decades.¹⁻⁸ When severe

it has a high risk of mortality, but few studies have quantified mortality after hospital admission in defined populations and time periods.⁴⁻¹⁰ We investigated long term trends in the incidence of acute pancreatitis with admission to hospital and in mortality after admission.

Methods

Study population

We used the Oxford record linkage study, which comprises anonymised hospital statistical records linked to death certificate data (population of 0.35 million from 1963, 0.9 million from 1968, 1.8 million from 1975, and 2.5 million from 1987). We selected all admissions for acute pancreatitis as the principal diagnosis recorded in the database. We identified each patient's first admission and any death that followed it within 365 days. At the end of each 365 day period we included any subsequent admission for a "new" period of one year follow up. The database covered admissions from 1963 to 31 March 1998 with linkage to death certificates to 31 March 1999.

Statistical methods

We used admissions for acute pancreatitis as the numerator and the total resident population in the area covered by the data as the denominator to calculate incidence rates for admission to hospital. We used the direct method and the standard European population to standardise trends in incidence rates. We used admissions for acute pancreatitis as the denominator and deaths from any cause after admission as the numerator to calculate case fatality rates. Our standard population to standardise trends in case fatality rates directly was the total population of patients admitted for acute pancreatitis from 1963 to 1998. We used the indirect method, applying the age and sex specific



Additional figures A and B and table A are on bmj.com