PERGAMON

European Journal of Cancer 39 (2003) 2278–2287

European Journal of Cancer

www.ejconline.com

Review

Randomised trials comparing chemotherapy regimens for advanced non-small cell lung cancer: biases and evolution over time

J.P.A. Ioannidis[a,c], A. Polycarpou[a], C. Ntais[a], N. Pavlidis[b,*]

[a]*Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece*

[b]*Division of Medical Oncology, University of Ioannina School of Medicine, Ioannina 45110, Greece*

[c]*Division of Clinical Care Research, Department of Medicine, Tufts University School of Medicine, Boston, MA, USA*

Received 24 April 2003; accepted 6 May 2003

Abstract

We systematically evaluated the evidence from randomised trials comparing various chemotherapy regimens for advanced non-small cell lung cancer. Across 254 eligible trials (42 661 patients), no regimens were compared in >6 studies. Twenty-six trials (10%) found statistically significant differences in survival between the compared arms. Only five reported the randomisation mode, and four reported adequate allocation concealment; nine performed unaccounted interim analyses. Statistical significance was more common in larger ($P=0.003$), more recent studies ($P=0.031$), and trials from countries with only one published eligible study ($P=0.008$). Increased reported median survival was independently associated with platinum and/or taxane and combination regimens, but also with the year of publication, smaller sample size, and larger representation of non-stage IV patients and patients with a better performance status. The proportion of enrolled patients with a performance status of 2 or worse decreased significantly over time (12.9% per decade, $P<0.001$). Randomised evidence in this field is fragmented and subject to considerable selection biases.

Keywords: Lung cancer; Non-small cell; Advanced stage; Meta-analysis; Bias; Randomised controlled trials; Survival; Performance status; Selection bias; Platinum; Taxane

1. Introduction

Chemotherapy may prolong survival in advanced, inoperable non-small cell lung cancer [1], although its effects are far less impressive than in small-cell lung cancer [2]. A meta-analysis of individual patient data [1] showed a small, but statistically significant, survival benefit over best supportive care, especially for platinum-containing regimens. Nevertheless, the overall prognosis of these patients remains poor. Much clinical effort has been devoted to identifying better chemotherapeutic regimens and many trials have been performed comparing different regimens. An evaluation of 33 phase III trials launched in North America between 1973 and 1994 ($n=8434$ patients) reached sobering conclusions [3]. Only five trials showed statistically significant survival benefits, all of less than 2 months' improvement in median survival. However, this analysis included only four trials conducted after 1990.

Several new trials continue to be performed and new regimens are being investigated in this field, including a rapidly increasing literature on taxanes (paclitaxel and docetaxel) in more recent years [4]. Moreover, the complete clinical experience of randomised trials in this setting includes many more investigations. Much research is conducted outside America and phase II and II/III trials far outnumber large phase III trials. Therefore, we decided to perform a systematic overview of all randomised clinical trials in patients with advanced non-small cell lung cancer. We systematised the available information and examined whether there is evidence for the superior efficacy of specific regimens against others, and whether reported significant survival differences are reliable or affected by biases. This information would

* Corresponding author. Tel. +30-26510-99394; fax: +30-26510-97505.

E-mail address: npavlid@cc.uoi.gr (N. Pavlidis).

0959-8049/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0959-8049(03)00571-9

provide insights for improving clinical practice and for guiding future clinical research in this field.

2. Patients and methods

2.1. Search strategy and eligibility criteria

We searched MEDLINE, EMBASE, and the Central Registry of Controlled Trials of the Cochrane Library [5] until November 2002. The search strategy used 'lung cancer' and 'chemotherapy' with an array of terms suggestive of randomised controlled trials, as recommended by the Cochrane Collaboration [5]. The full strategy is available upon request. In addition, we identified all previous systematic reviews and meta-analyses of randomised trials in this field and perused the references of retrieved articles. Cross-searches were performed in MEDLINE using the names of investigators who were lead authors in at least one eligible trial. Finally, we hand-searched several years of the volumes of journals that had the highest number of electronically identified trials [6].

We considered all randomised controlled trials that compared at least two arms of different chemotherapy regimens in patients with advanced non-small cell lung cancer (stage IIIB or IV, unresectable or recurrent). We excluded non-randomised trials and pseudo-randomised trials with alternate allocation of subjects. We excluded trials limited to non-advanced disease and trials limited to small-cell lung cancer. However, trials were eligible regardless of whether they might also have included some patients with non-advanced disease or other malignancies; deviations were noted and the number of strictly eligible patients was also recorded. We excluded trials comparing radiotherapy regimens, but retained trials when radiotherapy was the same in all of the compared arms and the difference between arms pertained to the chemotherapeutic regimens. We excluded trials where the compared arms used the same chemotherapy regimen, but differed in the non-chemotherapeutic agents used (e.g. immunomodulators, growth factors, cytokines or verapamil). We accepted randomised trials comparing different dosing schemes and schedules of the same agent or combination of agents. We excluded comparisons of chemotherapy against best supportive care without chemotherapy. Trials with three or more arms were retained if at least two arms addressed an eligible comparison; non-eligible arms were excluded from further analyses.

Trial reports were scrutinised to identify potential duplication and overlap. In cases of overlap or duplicate reports, only the main report with the maximal information was retained. Interim analyses were retained, if no further final report had been published. We only considered studies published in journals. Meeting abstracts were excluded. There was no language restriction.

2.2. Data extraction

From each eligible trial report, we recorded the authors, publication year, journal, sample size (total, eligible, per arm), regimens compared, type of chemotherapy (based on whether platinum, taxane, both, or neither was used), country(ies) of the investigators, study population eligibility criteria, and the percentage of patients with performance status 2 or worse (Karnofsky score 70 or less) and stage IV disease per arm. Furthermore, we recorded the median survival per arm, and whether there was any statistically significant difference in median survival between compared arms ($P<0.05$). When several different analyses were reported with different levels of statistical significance, we preferred the log-rank test results over other statistics. In trials with >2 arms, where statistically significant differences in survival were reported, we estimated whether statistical significance would still be claimed in a log-rank test comparing all arms, using the presented information from Kaplan–Meier curves, tables, text and pair-wise comparisons [7]. For trials with statistically significant survival differences, we recorded qualitative parameters including masking, adequate reporting of the mode of randomisation, allocation concealment and use of planned or unplanned interim analyses.

2.3. Statistical analyses

We classified regimens and comparisons according to the exact chemotherapy being used and according to broad categories defined by the use or not of platinum and/or taxanes; and according to whether monotherapy or at least two agents were involved. We evaluated using the Mann–Whitney U test and Fisher's exact test whether the types of regimens compared, trial sample size, year of publication and origin were associated with the chance of finding a statistically significant survival difference.

Excluding trials with a mixture of patients with other malignancies or earlier stage non-small cell lung cancer and trials where radiotherapy was part of the regimens, we evaluated whether the reported median survival was related to the type of regimen (containing platinum, taxane, or both), number of drugs (monotherapy versus combination of at least two agents), year of publication, sample size of study arm, performance status, disease stage and country of origin. We used analysis of variance and linear regression. Multivariate regression models were also evaluated starting from variables with $P<0.2$ in univariate analyses and dropping non-significant variables with backward elimination. We report unweighted regressions and regressions weighted by the sample size of each study arm.

Analyses were conducted in Statistical Package for the Social Sciences (SPSS) 11.0. *P* values are two-tailed.

3. Results

3.1. Eligible trials

Of 350 potentially eligible trial reports, we excluded 49 that were ineligible upon full-text scrutiny, three that could not be retrieved in full text and where abstracts did not provide even the total sample size, and 44 duplicate/preliminary presentations. Thus, 254 independent trials were retained (Table 1). The number of trials per year increased over time. The median sample size was 114, but only one trial had over 1000 subjects ($n = 1207$, a four-arm trial) and the largest study arm had 309 subjects. A total of 42 661 subjects were randomised across 594 arms. Six control arms where no chemotherapy was given ($n = 338$) were not analysed, leaving 588 eligible trial arms ($n = 42\,323$). In the large majority of reports, no previous chemotherapy had been given and a few early trials had included also some patients with non-advanced disease or other malignancies. In almost one-third of the trials, only US investigators were involved, while there was also a large number of Italian and Japanese trials. Trials with multinational authorship accounted for 18.9%.

Table 1
Characteristics of the eligible trials

Characteristic	
Year of publication	
1971–1980	34
1981–1990	91
1991–2000	98
2001–2002	31
Total sample size, median (IQR)	114 (68–207)
Eligible subjects, median (IQR)	106 (64–210)[a]
Number of eligible arms	
Two	197
Three	45
Four or more	12
Previous chemotherapy given, *n* (%)	26 (10.2)[b]
Including patients with non-advanced disease, *n* (%)	30 (11.8)
Including patients with other malignancies	
Small cell lung cancer, *n* (%)	18 (7.1)
Non-lung cancer, *n* (%)	8 (3.1)
Radiotherapy also involved, *n* (%)	8 (3.1)
Countries involved[c]	
USA	73
Multiple countries	48
Italy	42
Japan	28
Greece	8
China	8
Spain	7
United Kingdom	7
France	6
Other/unknown	27

IQR, interquartile range; USA, United States of America.

[a] Data available on 240 of the 254 trials.

[b] Only three trials included exclusively patients who had received previous chemotherapy.

[c] Based on affiliations of authors and investigators; it was not always possible to ascertain whether enrollment was conducted strictly in the same countries.

3.2. Compared chemotherapy regimens

Only seven monotherapies and 14 combinations of different agents had been tested in at least five trial arms each. These 21 regimens accounted for only 271 of the 588 arms (Table 2). Moreover, the doses and schedules of the same regimen often differed across trials. The most commonly assessed regimen, cisplatin + etoposide, was utilised in 35 trials (41 arms). Five trials (10 arms) simply compared various schedules and doses of this combination. In 31 trials, this combination was compared against 32 other different regimens in 44 possible comparisons. A comparison against cisplatin + vindesine was performed in five trials and a comparison against cisplatin + mitomycin + vinblastine in four others. The total number of subjects in these contrasts amounted to only 695 and 662 patients, respectively. There were two comparisons of cisplatin + etoposide against six different regimens and a single comparison against another 23 regimens.

Table 2
The 21 most commonly used regimens in the eligible trials

Regimen	Number of arms
Cisplatin + etoposide	41
Cisplatin + vindesine	35
Cisplatin + mitomycin C + vindesine	23
Carboplatin + paclitaxel	17
Cisplatin + cyclophosphamide + doxorubicin	16
Cisplatin	15
Cisplatin + gemcitabine	15
Cisplatin + vinorelbine	12
Cisplatin + mitomycin C + vinblastine	11
Cyclophosphamide	10
Cisplatin + paclitaxel	9
Vindesine	9
Carboplatin + etoposide	8
Cisplatin + vinblastine	8
Cyclophosphamide + methotrexate + doxorubicin + lomustine	7
Vinorelbine	7
Carboplatin	6
Cisplatin + ifosfamide + mitomycin C	6
Cispatin + ifosfamide + vindesine	6
Gemcitabine	5
Paclitaxel	5

The only comparison that occurred at least six times was cisplatin + vindesine versus cisplatin + mitomycin + vindesine, but all six comparisons accounted in total for only 500 patients. The only comparison that was evaluated in a total of over 1000 randomised patients was carboplatin + paclitaxel versus cisplatin + paclitaxel ($n = 1243$ across two phase III trials and one small phase II trial). One phase III trial found a significant survival prolongation with the cisplatin combination (median 9.8 versus 8.2 months), while the other phase III trial found the opposite trend (7.8 versus 8.1 months).

In terms of broad categories, 225 arms used regimens without platinum or taxanes, 313 platinum-based regimens without taxanes, 35 regimens with both platinum and taxanes, and 15 regimens with taxanes, but without platinum. Regimens with neither platinum nor taxanes are currently disappearing, while combinations of platinum and taxanes recently account for over a third of the tested regimens (Table 3). Seventy-one trials involved only comparisons between regimens without platinum or taxanes, 44 involved a comparison of a platinum-based non-taxane regimen against a regimen without platinum or taxane, 111 involved only comparisons of platinum-based regimens without taxanes, and 28 involved taxanes. Fifty trials involved at least one comparison of monotherapy against combinations.

3.3. Trials with statistically significant survival differences

Thirty-one trials claimed statistically significant differences in survival between the compared arms. In 5 cases, this involved only one of three or more pair-wise comparisons in multi-arm trials without significance in the overall comparison across all arms. Twenty-six trials (10%) had overall statistically significant survival differences (Table 4) [8–33]. None of them was blinded, only five (19%) described an appropriate mode of randomisation [8,18,20,25,32] and only four (15%) described an appropriate mode of allocation concealment [8,15,25,32]. Three were stopped early in interim analyses [8,22,24], including an unplanned one (the primary investigator left the institution) [24]. Another trial found significant survival differences only when follow-up was updated beyond the original analysis [32]. Finally, at least five other trials reported 1–2 additional interim analyses [9,13,17,25,33], and at least one was acknowledged to be unplanned [25]. No trial adjusted results for interim analyses.

Four trials involved only comparisons of non-platinum, non-taxane regimens [8–11], 12 included only comparisons of platinum-based, non-taxane regimens [12–23], eight included a comparison of a platinum-based, non-taxane regimen against non-platinum, non-taxane drugs [24–31], and two involved taxanes [32,33], although the compared regimens used the same taxane in one of them [32]. The percentage of statistically significant survival differences in these four trial categories was 6% (4/71), 11% (12/111), 18% (8/44) and 7% (2/28), respectively ($P = 0.20$). There were nine statistically significant survival differences across 50 trials where a monotherapy was compared against a combination(s) of agents versus 17 such differences across all other 204 trials (18% versus 8%, $P = 0.065$).

Trials with statistically significant results had a larger sample size ($P = 0.003$) and a later publication year ($P = 0.031$) on average than other trials. However, excluding 51 trials with sample sizes up to 60 patients, the differences became non-significant ($P = 0.18$ and $P = 0.17$, respectively).

Statistically significant survival differences were seen in 9/48 trials with a multinational authorship versus 17/206 other trials ($P = 0.059$). Six countries (Germany, Romania, South Africa, Sweden, Turkey, Yugoslavia) published only one single-country eligible trial. Three of these (Romania, Turkey, and Yugoslavia) claimed statistically significant survival differences. The difference against the proportion of statistically significant results from other trials with a single-country authorship was impressive (50% versus 7%, $P = 0.008$). All 3 cases involved comparisons between platinum-based, non-taxane regimens. Romania and Turkey were not involved in any published multinational studies.

The absolute difference between arms in median survival exceeded 3 months in nine trials. This included four trials comparing platinum-based regimens versus non-platinum chemotherapy, and five involving comparisons of platinum-based regimens. Only one trial with a sample size exceeding 160 patients [20] showed a statistically significant survival benefit exceeding 3 months.

3.4. Parameters affecting median survival

For this analysis, 196 trials with 462 arms were eligible and median survival per arm was available for 160 trials (374 arms). Mean median survival was 7.10 (standard deviation (S.D.) 2.32) months. Median survivals exceeding 12 months were recorded in seven arms.

Table 3
Regimens used in different decades (% in decade)

	NP/NT	P/NT	P/T	NP/T
1971–1980	73 (94)	5 (6)	0	0
1981–1990	95 (44)	121 (56)	0	0
1991–2000	49 (22)	152 (69)	8 (4)	10 (5)
2001–2002	8 (11)	35 (47)	27 (36)	5 (7)
Total	225	313	35	15

P, platinum; T, taxane; NP, no platinum; NT, no taxane.

Table 4
Trials with formally statistically significant differences in survival across compared arms

Author, year	N (eligible)[a]	Countries involved	Regimens compared	N_{arm}[a]	S_{median}[b]
Frasci, 2000	125 (125)	Italy	Gemcitabine + vinorelbine	60	6.7
			Vinorelbine	60	4.2
Crawford, 1996	216 (216)	Canada, USA	Vinorelbine	143	6.9
			Fluorouracil + leucovorin	68	5.1
Gatzemeier, 1991	184 (184)	Germany, Switzerland	Lonidamine	64	4.8
			Mitomycin C + vindesine	60	6.4
			Mitomycin C + lonidamine + vindesine	60	7.3
Green, 1990	178 (178)	USA, Canada	CisPl + cytarabine	79	6.0
			CisPl + vinblastine	76	7.2
Grigorescu, 2002	198 (187)	Romania	CarboPl + gemcitabine	95	9.3
			CisPl + vinblastine	92	7.6
Jelic, 2001	221 (221)	Yugoslavia	CisPl + mitomycin C + vindesine	114	8.0
			CarboPl + mitomycin C + vindesine	107	9.6
Ricci, 2000	82 (82)	Italy	CisPl + gemcitabine (schedule 1)	42	10.0
			CisPl + gemcitabine (schedule 2)	40	17.0
Sandler, 2000	522 (522)	Canada, Finland, Germany, UK, USA	CisPl + gemcitabine	260	9.1
			CisPl	262	7.6
von Pawel, 2000	446 (446)	Australia, Belgium, Canada, UK Germany, France, Sweden, USA	CisPl + tirapazamine	223	8.0
			CisPl	223	6.4
Wozniak, 1998	432 (432)	USA	CisPl	209	6.0
			CisPl + vinorelbine	206	8.0
Ginopoulos, 1997	85 (85)	Greece	CisPl + mitomycin C + vindesine	43	9.7
			CisPl + eto	42	6.9
Ianniello, 1996	158 (158)	Italy	CisPl + epirubicin + lonidamine + vindesine	78	11.0
			CisPl + epirubicin + vindesine	80	7.6
Crino, 1995	393 (393)	Italy	CisPl + eto	130	5.8
			CisPl + mitomycin C + vindesine	130	9.6
			CisPl + mitomycin C + ifosfamide	133	8.1
Erkisi, 1995	77 (77)	Turkey	CisPl + eto + ifosfamide	39	14.9
			CisPl + eto	35	9.1
Crino, 1990	156 (156)	Italy	CisPl	24	4.1
			CisPl + eto	69	9.7
			CisPl + eto + mitomycin C	57	8.1
Rapp, 1988	101 (101)	Canada	CisPl + vindesine	49	7.9
			CisPl + cyc + Adria	52	5.2
Jeremic, 1997	120 (120)	Japan, Yugoslavia	Eto	59	5.0
			CarboPl + eto	58	9.0
Le Chevalier, 1994	612 (612)	Belgium, France, Germany, Greece, Italy, Portugal, Spain, Switzerland	CisPl + vinorelbine	206	9.2
			CisPl + vindesine	200	7.4
			Vinorelbine	206	7.1
Veronesi, 1988	136 (?)	Italy	Adria + cyc + methotrexate + procarbazine	?	5.1
			CisPl + eto	?	6.8
Ohta, 1988	119 (117)	Japan	Adria + CisPl + cyc + mitomycin C	56	9.5
			Cytarabine + mitomycin C + tegafur	61	6.0
Miller, 1986	483 (483)	USA	5-Fluorouracil + mitomycin C + vincristine	158	4.6
			CisPl + cyc + Adria	166	5.5
			Alternating regimens listed above	159	5.3
Elliott, 1984	105 (105)	UK	Vindesine	54	4.0
			CisPl + vindesine	51	11.0
Fuks, 1983	68 (68)	Canada, USA	CisPl + cyc + adria + eto	38	8.0
			Cyc + adria + eto	30	5.2
Eagan, 1979	72 (72)	USA	Adria + cisPl + cyc + rad	34	16.6
			Adria + cyc + DTIC + rad	34	7.1
Rosell, 2002	618 (618)	15 European countries, Israel, USA	CarboPl + paclitaxel	309	8.2
			CisPl + paclitaxel	309	9.8
Comella, 2001	360 (360)	Italy	CisPl + gemcitabine	112	8.8
			CisPl + gemcitabine + vinorelbine	117	11.8
			CisPl + gemcitabine + paclitaxel	114	11.8

Adria, adriamycin (doxorubicin); CarboPl, carboplatin; CisPl, cisplatin; cyc, cyclophosphamide; eto, etoposide; rad, radiotherapy; DTIC, dimethyltriazenoimidazole carboxamide.

[a] N refers to the total number of randomised subjects, while the number in parentheses refers to patients with stage IIIB or IV non-small cell lung cancer. The number of subjects per arm, N_{arm}, pertains to the eligible patients entered in the analysis of the trial results. As shown, in 7 cases, the analysed patients are fewer than the total number randomised.

[b] S_{median} refers to the median survival, as estimated from Kaplan–Meier plots.

In unweighted analyses, the median survival was clearly different across classes of chemotherapy ($P<0.001$) with significantly shorter median survival when either platinum nor taxanes were included (mean 5.49 (S.D. 1.84) months based on 119 available arms). Differences between regimens including platinum, taxane, or both were small (mean (SD): 7.73 (2.11), 7.83 (2.25), and 8.77 (2.26) months based on 217, 11, and 27 arms, respectively). Combinations were associated with 1.84 months longer median survival (S.D. 0.30, $P<0.001$) than monotherapies. Median survival increased 1.72 (S.D. 0.15) months ($P<0.001$) per decade. For each 50 patients in the study arm, the median survival seemed to increase by 0.23 (S.D. 0.10) months ($P=0.020$). Median survival decreased by 0.29 months (S.D. 0.06, $P<0.001$) per each 10% increase in the proportion of subjects with a performance status of 2 or worse and 0.42 months (S.D. 0.06, $P<0.001$) for each 10% increase in the proportion of subjects with stage IV disease. There was also a shorter median survival in the multinational trials (−1.31 (S.D. 0.27) months, $P<0.001$) and a trend for a longer reported survival in trials from countries that had published only one trial (+1.35 (S.D. 0.89) months, $P=0.13$). Arm sample size was highly significantly correlated ($P<0.001$) with the year of publication ($r=0.34$), multinational authorship ($r=0.18$), better performance status ($r=0.24$) and use of platinum and/or taxane ($r=0.25$) and to a lesser extent with a higher proportion of stage IV disease ($r=0.16$, $P=0.003$) and the use of combination regimens ($r=0.10$, $P=0.045$).

Regardless of the year of publication, regimens that include neither platinum nor taxanes always showed an inferior survival compared with the other regimens (Fig. 1). Monotherapies on average reached a median survival in the range of 6 months. Over the last 25 years, the reported median survival improved by approximately 4 months, but within the combination regimens including platinum and/or taxanes, the recorded average improvement in the same time period was approximately 2 months. However, there was a strong correlation between the year of publication and the proportion of subjects with a poor performance status. In an unweighted regression, the proportion of enrolled subjects with a performance status of 2 or worse decreased by 10.9% per decade (S.D. 1.22, $P<0.001$, Fig. 2). The effect was similar in the weighted regression (12.9%, S.D. 1.07, $P<0.001$).

In multivariate modelling (Table 5), use of platinum and/or taxanes, combination of agents, performance status, proportion of subjects with stage IV disease, and year of publication were highly significant independent predictors of the median survival, in both the weighted and unweighted analyses. The adjusted effect of sample size was inversed with a modest decrease in reported survival with larger sample sizes. An independent effect of the specific combination of platinum and taxanes was discerned in the weighted analyses. Unweighted and

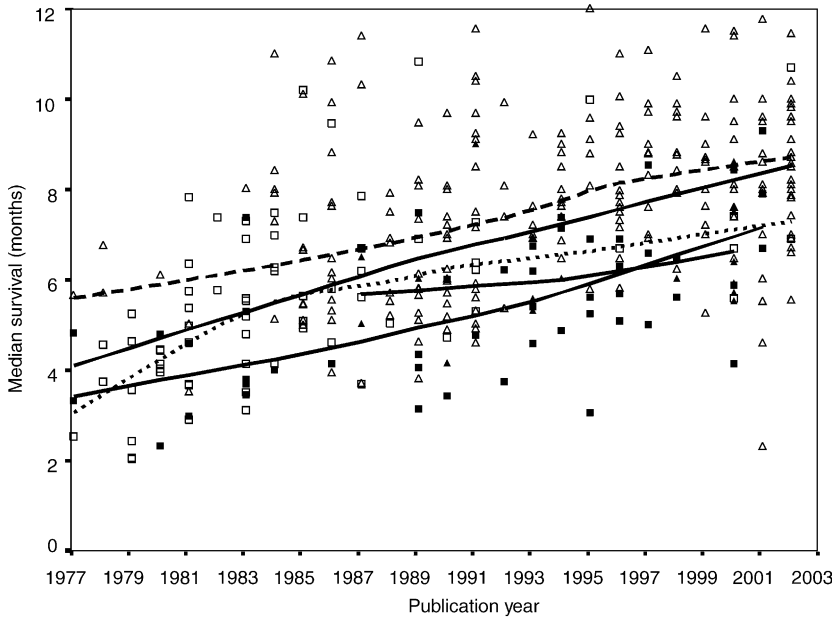

Fig. 1. The median survival per arm is plotted as a function of the year of publication of each trial. Triangles represent arms where platinum and/or a taxane is used, while squares represent trial arms where neither platinum nor taxanes are involved. Filled markers represent monotherapies, while open markers represent combinations of at least two agents. Also shown are best-fit lowess splines for all data (thick continuous line extending from 4 months in 1977 to 8 months in 2002) and separately for the four resulting subgroups (continuous lines for monotherapies, discontinuous lines for combinations). Seven arms with median survival exceeding 12 months are not shown, but are used in the calculations.

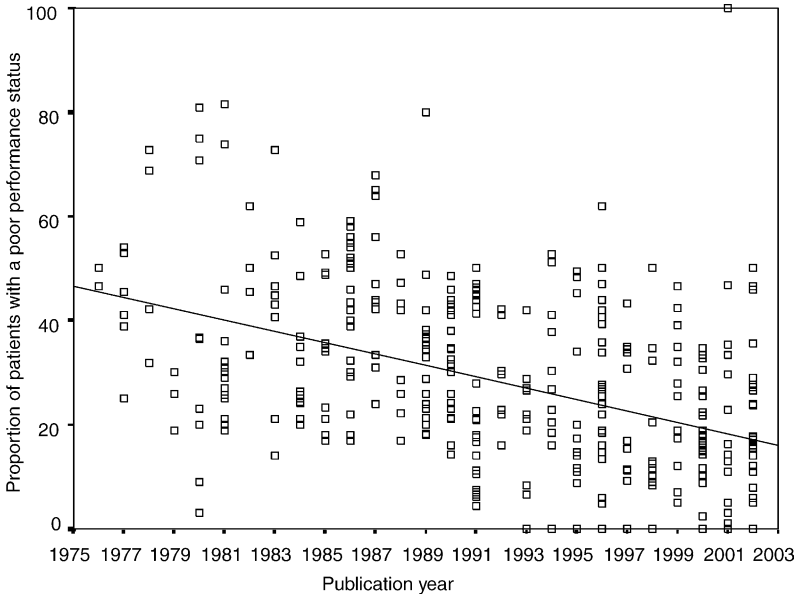

Fig. 2. Proportion of subjects in each study arm with a performance status of 2 or worse (or Karnofsky of 70 or worse) as a function of the year of publication of the trial.

weighted models explained 47 and 56% of the variability, respectively.

4. Discussion

Despite the availability of over 250 published trials on chemotherapy regimens for advanced non-small cell lung cancer, the evidence remains extremely fragmented. There is a large variability in the regimens used and comparisons thereof. With a total of over 40 000 randomised patients, with one exception, no regimens have been compared for their impact on survival on at least 1000 patients and single comparisons of two study arms have never involved at least 700 patients. Trials conducted to date are underpowered to detect plausible treatment effects. For a hazard ratio of 1.2, in order to have 80% power at $\alpha = 0.05$, 950 subjects must be followed until death, i.e. over 1000 patients should be randomised in two compared arms [34]. This has never been performed in single trials to date.

In search of novelty and under corporate pressure, investigators may improvise with selecting new contrasts for new trials, instead of comparatively evaluating and validating previously assessed regimens. Innovation is not necessarily bad. It allows the generation of some useful data in a randomised, controlled setting for an expanding array of agents and combinations. Safety and preliminary efficacy data are certainly welcome. However, this fragmented approach is unlikely to provide definitive evidence about the superiority of specific regimens in terms of survival benefits.

We suspect that chance findings (type I errors) are equally or even more problematic than type II errors in this field. We recorded 26 trials where a formally statistically significant survival difference had been observed. This represents 10% of the trial reports. One might expect by chance up to a 5% occurrence of statistically significant survival differences. With the exception of trials comparing platinum-based versus non-platinum regimens and trials comparing combinations of agents against monotherapies, the percentage of significant findings was not far from 5%. Several of these 26 'significant' findings may have been due to chance alone. The phenomenon of spurious differences in survival has been previously described empirically also in other medical domains [35]. Of note, almost all the statistically significant benefits exceeding 3 months gain in median survival were recorded in relatively small trials, further suggesting that these estimates are probably inflated by chance.

We also found considerable evidence for biases in this field. Lack of double-blinding, and no reporting of the mode of randomisation and of the allocation concealment have been associated with spuriously inflated treatment effects in other medical domains [36,37]. The trials analysed here scored very poorly in these aspects. Of course blinding is not convenient in this setting and perhaps comparative survival may not be as easily affected by methodological shortcomings. Reporting does not always reflect the actual study design [38] and methodological shortcomings have not been associated with inflated outcomes in all medical domains [39]. Nevertheless, improved study design and reporting

Table 5
Parameters potentially affecting the reported median survival: multivariate models

Parameters	Unweighted analysis		Weighted analysis[a]	
	Change in median survival (S.D.) months	*P* value	Change in median survival (S.D.) months	*P* value
Chemotherapy regimen with platinum and/or taxane	+0.76 (0.27)	0.005	+0.66 (0.24)	0.007
Combination of at least two agents	+1.41 (0.28)	<0.001	+1.17 (0.24)	<0.001
Combination of platinum and taxane	NS	NS	+0.58 (0.26)	0.029
Publication year (per decade)	+1.31 (0.17)	<0.001	+1.33 (0.16)	<0.001
Sample size of the study arm (per 50 patients)	−0.17 (0.08)	0.035	−0.18 (0.05)	0.003
Country with only one published trial	NS	NS	NS	NS
Multinational authorship	NS	NS	NS	NS
Performance status (per 10% increase of 2 or worse)	−0.23 (0.05)	<0.001	−0.23 (0.05)	<0.001
Stage IV disease (per 10% increase)	−0.36 (0.05)	<0.001	−0.37 (0.04)	<0.001

S.D., standard deviation; NS, not selected in the multivariate model (non significant).
[a] Weighted by the sample size of the study arm.

would be useful in this field. Interim analyses are also very common in these trials. This is appropriate in selected cases, but early stopping can lead to inflated estimates of treatment effects [40] and more common chance findings [41].

The country of origin was a strong determinant of reaching statistically significant results. Trials of multinational authorship are larger, so an association with significant results is not surprising. However, significant results were most strongly associated not with multinational origin, but with countries that had contributed only one single-country trial to the international literature. A tower of Babel bias may be operating [42]. In some countries, trials with statistically significant, favourable results may be selectively published in the international literature, while trials with less favourable results may appear in local non-indexed publications or may remain unpublished [43]. We cannot probe the exact depth of the potential tower of Babel and publication biases here [44]. However, given that non-small cell lung cancer is a very common cancer worldwide and chemotherapy trials are easy to conduct, we suspect that a sizeable proportion of the relevant literature from non-English speaking countries remains unpublished and/or largely inaccessible. This may not be a problem for large trials conducted by well-established multicentre groups [3], but the total literature may give misleading impressions.

Trends in observed median survival should be interpreted with caution since they are based on data from single arms, thus the benefit of randomisation is lost and single-arm data represent observational cohorts [45]. Even though we tried to retain the most homogeneous study populations for these analyses, changes in medical management and supportive care over time are difficult to model. The improvement over time correlated in part to the introduction of platinum and combination regimens. Combinations including specifically platinum and taxanes may confer an additional, independent benefit, but the estimate of this benefit is very small and highly susceptible to potential bias. Comparator non-platinum agents included in early treatment regimens might have even had a deleterious effect on survival, as has been shown for several alkylating agents when compared against best supportive care [1]. Moreover, we clearly documented a strong selection shift with a decrease in the proportion of enrolled patients with a poor performance status in randomised trials over time. Performance status is a very strong predictor of survival and may account at least in part for the observed improved survival over time. Given the high correlation between the year of publication, parameters representing treatment interventions and parameters representing bias, multicollinearity may be a problem and adjusted estimates may be misleading. Epidemiological evidence [3] suggests that median survival in patients with advanced non-small cell lung cancer in the US population has improved by only a few weeks over the last 3 decades. Thus, the observed improvement in median survival in clinical trials may be highly spurious.

Several paradigm shifts are required in clinical research on advanced non-small cell lung cancer. Our analysis expands and updates the previous suggestion that treatment approaches in this setting carry sobering results [3] and adds concerns about prevalent biases in the accumulated evidence. Any survival gains with currently available regimens seem very modest at best [46]. Research priorities should probably focus on developing new treatment modalities and on preventive efforts

[47]. Performing hundreds of underpowered chemotherapy trials represents a poor, uncoordinated research investment. Larger, well-designed and adequately reported trials may be more definitive [48,49], but their targets should be carefully selected. Smaller trials are still useful, and careful meta-analyses [50,51] may yield some more conclusive results. However, given the extreme complexity and fragmentation of potential comparisons in this field, even meta-analyses are underpowered and selective inclusion of trials may generate spurious results. We need international collaboration, strategic planning of treatment assessments, and complete international registration of all trials conducted worldwide [52]. Otherwise, reported trial results are likely to remain fragmented or even misleading.

References

1. Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical trials. *Br Med J* 1995, **311**, 899–909.
2. Chute JP, Chen T, Feigal E, Simon R, Johnson BE. Twenty years of phase III trials for patients with extensive-stage small-cell lung cancer: perceptible progress. *J Clin Oncol* 1999, **17**, 1794–1801.
3. Breathnach OS, Freidlin B, Conley B, *et al*. Twenty-two years of phase III trials for patients with advanced non-small-cell lung cancer: sobering results. *J Clin Oncol* 2001, **19**, 1734–1742.
4. Crown J, O'Leary M. The taxanes: an update. *Lancet* 2000, **355**, 1176–1178.
5. www.cochrane.org
6. Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M. A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. *Stat Med* 2002, **21**, 1625–1634.
7. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*, 2nd ed. New York, Wiley, 2002.
8. Frasci G, Lorusso V, Panza N, *et al*. Gemcitabine plus vinorelbine versus vinorelbine alone in elderly patients with advanced non-small-cell lung cancer. *J Clin Oncol* 2000, **18**, 2529–2536.
9. Crawford J, O'Rourke M, Schiller JH, *et al*. Randomized trial of vinorelbine compared with fluorouracil plus leucovorin in patients with stage IV non-small-cell lung cancer. *J Clin Oncol* 1996, **14**, 2774–2784.
10. Gatzemeier U, Cavalli F, Haussinger K, *et al*. Phase III trial with and without lonidamine in non-small cell lung cancer. *Semin Oncol* 1991, **18**(Suppl. 4), 42–48.
11. Green MR, Kramar A, Schilsky R, *et al*. Platinum-based combination chemotherapy in advanced non-small cell lung cancer: a randomized phase II trial of the Cancer and Leukemia Group B. *Med Pediatr Oncol* 1990, **18**, 197–202.
12. Grigorescu AC, Draghici IN, Nitipir C, Gutulescu N, Corlan E. Gemcitabine (GEM) and carboplatin (CBDCA) versus cisplatin (CDDP) and vinblastine (VLB) in advanced non-small-cell lung cancer (NSCLC) stages III and IV: a phase III randomised trial. *Lung Cancer* 2002, **37**, 9–14.
13. Jelic S, Mitrovic L, Radosavljevic D, *et al*. Survival advantage for carboplatin substituting cisplatin in combination with vindesine and mitomycin C for stage IIIB and IV squamous-cell bronchogenic carcinoma: a randomized phase III study. *Lung Cancer* 2001, **34**, 1–13.
14. Ricci S, Antonuzzo A, Galli L, *et al*. A randomized study comparing two different schedules of administration of cisplatin in combination with gemcitabine in advanced nonsmall cell lung carcinoma. *Cancer* 2000, **89**, 1714–1719.
15. Sandler AB, Nemunaitis J, Denham C, *et al*. Phase III trial of gemcitabine plus cisplatin versus cisplatin alone in patients with locally advanced or metastatic non-small-cell lung cancer. *J Clin Oncol* 2000, **18**, 122–130.
16. von Pawel J, von Roemeling R, Gatzemeier U, *et al*. Tirapazamine plus cisplatin versus cisplatin in advanced non-small-cell lung cancer: a report of the international CATAPULT I study group. *J Clin Oncol* 2000, **18**, 1351–1359.
17. Wozniak AJ, Crowley JJ, Balcerzak SP, *et al*. Randomized trial comparing cisplatin with cisplatin plus vinorelbine in the treatment of advanced non-small-cell lung cancer: a Southwest Oncology Group study. *J Clin Oncol* 1998, **16**, 2459–2465.
18. Ginopoulos P, Spyropoulos K, Kardamakis D, *et al*. Advanced non-small cell lung cancer chemotherapy: a randomized trial of two active regimens (MVP and PE). *Cancer Lett* 1997, **119**, 241–247.
19. Ianniello GP, De Cataldis G, Comella P, *et al*. Cisplatin, epirubicin, and vindesine with or without lonidamine in the treatment of inoperable non-small-cell lung carcinoma: a multicenter randomized clinical trial. *Cancer* 1996, **78;**, 63–69.
20. Crino L, Clerici M, Figoli F, *et al*. Chemotherapy of advanced non-small-cell lung cancer: a comparison of three active regimens. A randomized trial of the Italian Oncology Group for Clinical Research (G.O.I.R.C.). *Ann Oncol* 1995, **6**, 347–353.
21. Erkisi M, Doran F, Burgut R, Kocabas A. A randomised trial of two cisplatin-containing chemotherapy regimens in patients with stage III-B and IV non-small cell lung cancer. *Lung Cancer* 1995, **12**, 237–246.
22. Crino L, Tonato M, Darwish S, *et al*. A randomized trial of three cisplatin-containing regimens in advanced non-small-cell lung cancer (NSCLC): a study of the Umbrian Lung Cancer Group. *Cancer Chemother Pharmacol* 1990, **26**, 52–56.
23. Rapp E, Pater JL, Willan A, *et al*. Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer—report of a Canadian multicenter randomized trial. *J Clin Oncol* 1988, **6**, 633–641.
24. Jeremic B, Shibamoto Y, Acimovic L, Milicic B, Milisavljevic S, Nikolic N. Prolonged administration of oral etoposide alone or with intravenous carboplatin in stage IV non-small cell lung cancer: a randomized trial. *Lung Cancer* 1997, **18**, 179–188.
25. Le Chevalier T, Brisgand D, Douillard J-Y, *et al*. Randomized study of vinorelbine and cisplatin versus vindesine and cisplatin versus vinorelbine alone in advanced non-small-cell lung cancer: results of a European multicenter trial including 612 patients. *J Clin Oncol* 1994, **12**, 360–367.
26. Veronesi A, Magri MD, Tirelli U, *et al*. Chemotherapy of advanced non-small cell lung cancer with cyclophosphamide, adriamycin, methotrexate, and procarbazine versus cisplatin and etoposide. A randomized study. *Am J Clin Oncol* 1988, **11**, 566–571.
27. Ohta M, Hara N, Ichikawa Y, *et al*. Benefits of cisplatin-based polychemotherapy in non-small cell bronchogenic carcinoma. Kyushu Lung Cancer Chemotherapy Study Group. *Gan To Kagaku Ryoho* 1988, **15**, 1901–1908.
28. Miller TP, Chen TT, Coltman Jr CA, *et al*. Effect of alternating combination chemotherapy on survival of ambulatory patients with metastatic large-cell and adenocarcinoma of the lung. A Southwest Oncology Group Study. *J Clin Oncol* 1986, **4**, 502–508.
29. Elliott JA, Ahmedzai S, Hole D, *et al*. Vindesine and cisplatin combination chemotherapy compared with vindesine as a single agent in the management of non-small cell lung cancer: a randomized study. *Eur J Cancer Clin Oncol* 1984, **20**, 1025–1032.

30. Fuks JZ, Aisner J, Van Echo DA, *et al*. Randomized study of cyclophosphamide, doxorubicin, and etoposide (VP16-213) with or without cisplatinum in non-small cell lung cancer. *J Clin Oncol* 1983, **1**, 295–301.
31. Eagan RT, Lee RE, Frytak S, *et al*. Randomized trial of thoracic irradiation plus combination chemotherapy for unresectable adenocarcinoma and large cell carcinoma of the lung. *Int J Radiat Oncol Biol Phys* 1979, **5**, 1401–1405.
32. Rosell R, Gatzmeier U, Betticher DC, *et al*. Phase III randomized trial comparing paclitaxel/carboplatin with paclitaxel/cisplatin in patients with advanced non-small-cell lung cancer: a cooperative multinational trial. *Ann Oncol* 2002, **13**, 1539–1549.
33. Comella P. Phase III trial of cisplatin/gemcitabine with or without vinorelbine or paclitaxel in advanced non-small cell lung cancer. *Semin Oncol* 2001, **28**(Suppl. 7), 7–10.
34. Piantadosi S. *Clinical Trials: A Methodologic Perspective*. New York, Wiley, 1997.
35. Ioannidis JP, Lau J. The impact of high-risk patients on the results of clinical trials. *J Clin Epidemiol* 1997, **50**, 1089–1098.
36. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995, **273**, 408–412.
37. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *Br Med J* 2001, **323**, 42–46.
38. Ioannidis JP, Lau J. Can quality of clinical trials and meta-analyses be quantified? *Lancet* 1998, **352**, 590–591.
39. Balk EM, Bonis PA, Moskowitz H, *et al*. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002, **287**, 2973–2982.
40. Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials* 1989, **10**(Suppl. 4), 209S–221S.
41. Lan KK, Rosenberger WF, Lachin JM. Use of spending functions for occasional or continuous monitoring of data in clinical trials. *Stat Med* 1993, **12**, 2219–2231.
42. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997, **350**, 326–329.
43. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990, **263**, 1385–1389.
44. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989, **84**, 374–380.
45. Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials* 2000, **21**, 488–497.
46. Ioannidis JPA, Pavlidis NA. Levels of absolute survival benefit for systemic therapies of advanced cancer: a call for standards. *Eur J Cancer* 2003, **39**, 1194–1198.
47. Bailar 3rd JC, Gornik HL. Cancer undefeated. *N Engl J Med* 1997, **336**, 1569–1574.
48. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984, **3**, 409–422.
49. Altman DG, Schulz KF, Moher D, *et al*. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001, **134**, 663–694.
50. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998, **351**, 123–127.
51. Clarke M, Stewart L, Pignon JP, Bijnens L. Individual patient data meta-analysis in cancer. *Br J Cancer* 1998, **77**, 2036–2044.
52. Jones TC. Call for a new approach to the process of clinical trials and drug registration. *Br Med J* 2001, **322**, 920–923.