Published in final edited form as:

J Biopharm Stat. 2016 ; 26(1): 141–149. doi:10.1080/10543406.2015.1092030.

Nine-year change in statistical design, profile, and success rates of Phase II oncology trials

Anastasia Ivanova^a, Barry Paul^b, Olga Marchenko^c, Guochen Song^c, Neerali Patel^d, and Stergios J. Moschos^b

^aDepartment of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

^bDepartment of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

^cQuintiles, Morrisville, North Carolina, USA

^dDepartment of Health Policy and Management, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Abstract

We investigated nine-year trends in statistical design and other features of Phase II oncology clinical trials published in 2005, 2010, and 2014 in five leading oncology journals: *Cancer*, *Clinical Cancer Research*, *Journal of Clinical Oncology*, *Annals of Oncology*, and *Lancet Oncology*. The features analyzed included cancer type, multicenter vs. single-institution, statistical design, primary endpoint, number of treatment arms, number of patients per treatment arm, whether or not statistical methods were well described, whether the drug was found effective based on rigorous statistical testing of the null hypothesis, and whether the drug was recommended for future studies.

Keywords

Fleming's design; oncology; Phase II trials; Simon's design; two-stage design

1. Introduction

Clinical trials are crucial in the development of safer and more effective cancer treatments, which may in part account for the decline in cancer-specific mortality over the last two decades (Siegel et al., 2015). Phase II oncology clinical trials are designed to determine whether an investigational treatment regimen has promising activity that warrants further clinical development in a Phase III trial, while more information is garnered regarding its toxicity profile than in a Phase I trial. Adherence to precise definitions about the null hypothesis using appropriate alpha and beta errors will support confident conclusions about

the primary endpoints tested; thus, a “successfully” conducted Phase II trial will potentially limit any considerations about performing *post hoc* analysis to explain outcomes.

Mariani and Marubini (2000) reviewed 309 Phase II cancer trials published between 1990 and 1996 and found that the overall quality of the statistical and methodological section was poor, since approximately 20% of the trials did not “well describe” statistical methods. Thezenas et al. (2004) reviewed Phase II trials published in 1995 and 2000 and restricted their analysis to studies with tumor response as the primary endpoint. Their primary goal was to investigate whether published papers described a statistical design or not, and if reported, which particular statistical design was used. We compared Phase II trials published in 2005, 2010, and 2014. This time interval has been associated with the development of novel anticancer treatments with diverse mechanisms of action, and a dramatic rise in the number of FDA approvals in oncology (Vera-Badillo et al., 2013). The goal of our study was to investigate the 9-year trend in success rates as measured by the proportion of trials that rejected the null hypothesis, and by the proportion of trials that recommended treatment for further investigation. We were also interested in whether the success rate was associated with a particular statistical methodology and whether the statistical design was explicitly described in the paper.

2. Selection of papers and data extraction

An electronic bibliographic database was utilized to identify leading oncology journals based on their reported impact factor. We searched PubMed and we limited our search to only Phase II clinical trials published in 2005, 2010, and 2014 in the following peer-reviewed journals: *Annals of Oncology* (AO), *Cancer* (CAN), *Clinical Cancer Research* (CCR), *Journal of Clinical Oncology* (JCO), and *Lancet Oncology* (LAN). We also searched the *New England Journal of Medicine* and *Lancet*, which occasionally report oncology trials, but we did not find any published Phase II clinical trial during each of these three years in the two journals. The search was conducted in each journal separately. For example, for Phase II studies published in the journal *Cancer* in 2005, we searched for articles using the following terms within Medical Subject Headings (MeSH) headings: (“Cancer”[Jour] AND Clinical Trial [ptyp] AND (“2005/01/01”[PDAT]: “2005/12/31”[PDAT])) AND (phase II[Title] OR phase 2[Title]). We removed articles that presented early results, described Phase II–III studies, and reported long-term follow-up or secondary analysis, because not all fields of interest were identifiable in these publications.

The following information was extracted from each article: publication year, article title, journal elements (name, volume, issue, number, pages), authors, country of origin (North America, Europe, Asia, Australia, and “rest of world”; in multinational studies the “country of origin” was recorded as the country of the first author), multicenter trial (yes or no), multinational trial (yes or no), cooperative oncology group trial (yes or no), cancer type, number of arms, and the total number of evaluable patients. With respect to the statistical aspects, the following information was extracted: primary endpoint, statistical design used, randomized design (yes or no), whether there was formal hypothesis testing, whether the null hypothesis was rejected, whether the therapy was recommended for future investigation, and whether statistical methods were “well described”. The primary outcome was recorded

as the outcome used in sample size justification (e.g. toxicity, overall survival (OS), progression-free survival (PFS), response rate (RR)). We defined statistical methods as “well described” both if the primary endpoint was explicitly defined and if there was justification of the sample size based on the primary endpoint.

A total of 347 published Phase II trials were selected from the following journals (2005, 2010, and 2014): CAN (35, 30, and 13), CCR (16, 20, and 3), JCO (63, 58, and 14), AO (26, 31, and 11), and LAN (1, 11, and 15). Considerably fewer Phase II trials that met our inclusion criteria were published in these journals in 2014 ($n = 56$) compared with that in 2005 ($n = 141$) and 2010 ($n = 150$). Of note, the number of Phase II trials published annually in these journals between 2006 and 2013 was comparable to those seen in 2005 and 2010.

Table 1 displays the proportion of trials by continent and features related to study structure. There were more multicenter and multinational trials published in 2014 compared to those in 2005 and 2010 (Table 1). With a handful of exceptions (e.g. lymphoma), the distribution of the cancer site was very similar during each of the three selected years (Table 2).

3. Primary endpoint and statistical design

Table 3 displays the primary endpoint used. In oncology, objective antitumor response in solid tumors is formally assessed using response evaluation criteria in solid tumors (RECIST), which were originally presented in 2000 (Therasse et al., 2000) and were subsequently refined in 2009 (Eisenhauer et al., 2009). Based on RECIST version 1.1 criteria, response is defined as: complete response (CR), the disappearance of all target lesions; partial response (PR), a 30% decrease of the longest diameter of all target lesions; progressive disease (PD), a 20% increase in the sum of the longest diameter of target lesions, taking as reference the smallest sum of the longest diameter recorded since the treatment started or the appearance of one or more new lesions. Responses that are not CR, PR, or PD are classified as stable disease (SD). For non-solid malignancies various International Working Groups have convened to establish criteria for CR, PR, SD, and PD across specific hematologic malignancies. Additional definitions that are frequently used to summarize antitumor responses are: RR, also called objective response or OR, defined as the sum of CR and PR; disease control rate (DC), defined as the sum of CR, PR, and SD; and clinical benefit (CB), defined as sustained DC (although there is no standardized definition of what length of time constitutes “sustained”). Other clinical endpoints include OS, defined as the time from investigational treatment initiation until death from any cause, PFS, defined as the time from investigational treatment initiation until objective tumor progression or death, and time-to-progression (TTP), defined as the time from investigational treatment initiation until objective tumor progression not including deaths.

While RR was the most frequently used primary endpoint in published clinical trials in 2005 and 2010, there was an increase in the use of OS/PFS/TTP in 2014. Approximately half of the trials that used PFS as the primary endpoint defined the primary outcome as the PFS probability within a certain finite time interval. Although the time interval varied considerably (range: 1 month, 2 years), the 6-month PFS rate was used most frequently. Similarly, in the trials that used OS as the primary endpoint, OS within a certain finite time

interval was used in approximately half of the trials, with OS at 1 year used most frequently (range: 6 months, 2 years). The advantage of using binary outcome instead of PFS or OS was the applicability of Simon's two-stage design (Simon, 1989). The PFS or OS probability within the specified interval was infrequently reported when results were presented and median survival was reported instead. One interesting example of dichotomizing PFS was the trial in Hainsworth et al. (2010) where the PFS rate at 6 weeks was used for futility stopping in Simon's design, and the PFS rate at 12 weeks was used for the final analysis.

Table 4 presents results for the study design. Many of the designs in Table 4 were recommended by the clinical trial design task force (Seymour et al., 2010) as appropriate Phase II study designs. In some articles, the type of the design used in a single-arm trial was not specified. If it was a two-stage design with a possibility of stopping for futility at the interim analysis based on a binary outcome, we recorded the design as being Simon's design (Simon, 1989). Simon's design remained the most frequently used two-stage design in Phase II oncology trials analyzed (40%). Increased use of Fleming's design (Fleming, 1982), a two-stage design with a possibility to stop for futility or efficacy after stage 1, was seen, from 2% in 2005 to 11% in 2014. Three percent of all trials used Bayesian designs. Of note, none of the seven articles where Bayesian designs were applied actually presented formal data analysis using the Bayesian methodology; instead, the data were summarized in a usual (frequentist) way. In addition, in many studies that used Bayesian methods the study design was not described probably due to its complexity. The usage of Gehan's design (Gehan, 1961) decreased from 6% in 2005 to 1% and 3% in 2010 and 2014, respectively. Three trials that were published in 2010 applied flexible designs, which allowed setting the sample size in the second stage according to the results observed in stage 1 (Bauer and Kohn, 1994; Chen and Ng, 1998). A two-stage design with stopping rules based on either response or toxicity (Bryant and Day, 1995) was used in three single-arm trials. One trial (Baselga et al., 2010) used an extension of Simon's design to ordinal outcomes (Lu et al., 2005), tumor response, and disease control. In another trial (Escudier et al., 2014) similar Simon's-like two-stage design for ordinal outcomes (Dent et al., 2001) was used. Two trials published in 2010 used the design by Sargent et al. (2001). One trial (Miller et al., 2014) used a two-stage design with stratification according to a prior therapy from London and Chang (2005). The triangular test (Bellissant et al., 1990) was used in one trial (Bernier-Chastagner et al., 2005). We also investigated the utilization of sequential stopping rules for toxicity and report the results separately in Ivanova et al. (2015).

In multi-arm trials where arms were not directly compared with each other, Simon's design in each arm was the most frequently applied design (42%), followed by a single-stage design (41%). In comparative multi-arm trials, a single-stage approach was the most frequently seen (70%). In many comparative multi-arm trials the "pick-the-winner" method that selects the best arm (Simon and Wittes, 1985) was used as opposed to classical hypothesis testing. The adaptive method of Bauer and Kohn (1994) was used in a comparative trial with two arms (Glass et al., 2014), and a Bayesian method with double criteria for noninferiority based on PFS (Neuenschwander et al., 2011) was used in a comparative two-arm trial described in Motzer et al. (2014).

Regarding statistical analysis methods, we saw the competing risks approach employed in a number of trials in hematologic oncology to account for a possible dependence of time-to-event outcomes, such as remission and death (Schmoor et al., 2013).

We analyzed the average number of patients per arm obtained as the total number of evaluable patients divided by the number of arms. The average number of patients per arm has been increasing; the median number of patients was 39 in 2005 compared to 45 in both 2010 and 2014. The proportion of trials with more than 60 patients per arm increased from 18% in 2005 to 30% in 2014.

4. The probability of finding a successful therapy

The proportion of manuscripts that do not explicitly describe the statistical design decreased slightly: 22 out of 141 articles (16%) published in 2005 compared to 17 out of 150 (11%) in 2010 and 7 out of 56 (12%) in 2014 (Table 5). The proportion of trials where the null hypothesis was formally tested has been increasing gradually from 74% in 2005 to 81% in 2010 and 93% in 2014. The proportion of trials rejecting the null hypothesis has been also increasing gradually from 38% in 2005 to 56% in 2014. Surprisingly, whether or not the null hypothesis was rejected was almost never mentioned in the abstract or even within the article itself. On most occasions we were able to establish whether or not the null hypothesis was rejected by comparing the decision rule in the design with the observed counts or by looking at reported confidence intervals. Sometimes we had to perform more sophisticated calculations, for example, in the case where Simon's two-stage design was used and the number of evaluable patients was not the same as specified in Simon's two-stage design.

Table 5 illustrates the relationship between rejection of the null hypothesis and whether the therapy was recommended for further investigation or not. In some trials, although the formal null hypothesis was not rejected, the drug was recommended for future investigation. On some occasions, this was based on the fact that the observed outcome was close to the rejection boundary of the null hypothesis. In other trials, the response in the primary endpoint (e.g. RR) was not impressive; however, the secondary endpoint (e.g. PFS or OS) was promising, and therefore the therapy was recommended for future research. In one trial, the null hypothesis regarding RR was rejected, but the therapy, a combination of weekly paclitaxel and gemcitabine, was not recommended due to high incidence of pulmonary toxicity (Li et al., 2005). We investigated how various factors, such as the primary endpoint and the number of patients per arm, affected the likelihood of rejecting the null hypothesis by fitting a logistic regression model. We did not find any significant association between rejection of the null hypothesis and any of the factors we considered.

One of the striking findings was that the probability of recommending a new therapy for future investigation was much higher in trials where statistical methods were not "well described". Of the 301 trials with "well described" statistical design, 206 (68%) of the articles recommended the therapy for future investigation. However, of the 46 trials with incomplete description of statistics, 41 (89%) of the articles recommended the therapy for future clinical development (Table 5). That is, the therapy was more likely to be recommended (p -value = 0.003) if the statistical method was not "well described" in the

article. The association remained significant (p -value = 0.005) after adjusting for the publication year and the number of patients per arm.

The association of several factors with whether or not the statistical method was “well described” is discussed in Table 6. The likelihood of statistical methods being “well described” is higher for publications originated from US-led compared with European-led clinical trials. These findings are similar to those in Thezenas et al. (2004); American-led trials and trials with more patients were more likely to describe statistical design and analysis in the corresponding publication. Thezenas et al. (2004) investigated the number of articles with explicitly stated statistical design and we looked at the number of articles where statistical methods were “well described”.

5. Discussion

We performed a comprehensive analysis of statistical design utilization and various other features associated with Phase II oncology clinical trials that have been published in peer-reviewed oncology journals in 2005, 2010, and 2014. This time interval was associated with dramatic changes in the clinical development of oncology treatments, such as new treatment types with different mechanisms of action than conventional cytotoxic therapies (e.g. immunotherapies, biologic agents, and highly selective small molecule inhibitors), more FDA-approved therapies, and demands for higher cost-effectiveness.

RR remains the preferred clinical endpoint to assess the clinical efficacy of an anticancer agent. However, PFS/TTP/OS are increasingly being used as clinically meaningful endpoints for efficacy, in particular for anticancer therapies that do not frequently induce objective antitumor responses (e.g. immunotherapies). Single-stage or Simon’s two-stage designs have been the most frequently used statistical designs in Phase II clinical trials. We found that more novel designs were being developed and, more importantly, implemented in oncology trials. One reason is the availability of easy-to-use software to design an oncology trial. Some of the designs for a single-arm Phase II oncology trial mentioned in the article, including Simon’s and Fleming’s designs and two-stage designs for ordinal outcomes, can be found at <http://cancer.unc.edu/biostatistics/program/ivanova/>.

Several interesting findings from our study highlight the impact of statistical design on decision making for further clinical development in oncology. We found that the proportion of trials where the null hypothesis was formally tested increased in 2010 and 2014 compared to 2005. This is an important finding for a number of reasons. First, there is an increasing trend over the last 10 years toward preliminary efficacy assessment of drugs in Phase I trials, in addition to safety. Therefore, more “promising” drugs from Phase I trials are being selected out for further (i.e. Phase II) clinical development. Second, we have better knowledge about the mechanism(s) of action of most anticancer drugs; this along with a better understanding of the biology of cancer has significantly contributed to more rational, and therefore more successful, clinical development. Third, go/no-go decisions for further clinical development of novel oncology treatments more strictly adhere to statistical results from Phase II clinical trials, because they are more robust (e.g. large numbers of patients per

arm) and due to competition from several other drugs with similar mechanisms of action under development.

Nevertheless, there is still a considerable number of articles with incomplete presentation of statistical design. We found that the corresponding therapies that are presented in these articles are more likely to be recommended. This underscores the importance of clear and careful planning of the trial and the significance of careful statistical designs in Phase II trials. In addition, somewhat unexpected, was the finding that articles almost never mentioned whether or not the null hypothesis was rejected. We recommend adherence to more rigorous reporting guidelines for description of statistical design and hypothesis testing by reviewers and editors in peer-reviewed journals.

The number of patients per arm has increased in trials published in 2014 articles. We postulate that these findings are related to the increasing number of clinical trials sponsored by pharmaceutical companies who wish to gain better confidence about the efficacy/toxicity profile of a particular regimen before commitment to costly Phase III (registrational) clinical trials. To this end, we are more frequently seeing pharma-sponsored Phase I clinical trials being amended in order to include multiple cancer-type-specific cohorts for efficacy testing, once the recommended Phase II dose (maximum tolerated dose or biologically effective dose) is established. By extending Phase I trials to include the desired disease-specific cohorts, Phase III studies can be directly planned if the efficacy signal is “promising”, just from Phase I data. This creates the need for rigorous statistical designs and reporting criteria for these Phase I/II studies with disease-specific cohorts.

Our investigation has several inherent limitations in the interpretation of results. We only included several leading oncology journals. For example, we did not include relevant publications from the journal *Blood*, a top hematologic malignancy journal, since hematologic malignancy studies were already well represented in our search. Additionally, clinical trials summarized in peer-reviewed oncology journals are only a small percentage of oncology trials that are conducted. For example, we did not report any clinical trials summarized in abstracts presented in meetings. Furthermore, publication bias can very well account for the higher number of articles that report on trials in which the null hypothesis is rejected and recommend further clinical development. We did not include Phase I/II studies in our search and therefore the trend of increasing number of Phase I/II trials may bias our 2014 results more than the earlier years. Finally, a lack of explicit description of the statistical design in the corresponding article may simply reflect the authors’ and journal editors’ style and prerogative.

Acknowledgments

The authors thank Anna Farrell and Donna Lague for editorial assistance. The authors thank an anonymous reviewer for helpful comments.

References

Baselga J, Gelmon KA, Verma S, Wardley A, Conte P, Miles D, Bianchi G, Cortes J, McNally VA, Ross GA, Fumoleau P, Gianni L. Phase II trial of pertuzumab and trastuzumab in patients with

- human epidermal growth factor receptor 2-positive metastatic breast cancer that progressed during prior trastuzumab therapy. *Journal of Clinical Oncology*. 2010; 28:1138–1144. [PubMed: 20124182]
- Bauer P, Kohne K. Evaluation of experiments with adaptive interim analyses. *Biometrics*. 1994; 50:1029–1041. [PubMed: 7786985]
- Bellissant E, Benichou J, Chastang C. Application of the triangular test to phase II cancer clinical trials. *Statistics in Medicine*. 1990; 9:907–917. [PubMed: 2218193]
- Bernier-Chastagner V, Grill J, Doz F, Bracard S, Gentet JC, Marie-Cardine A, Luporsi E, Margueritte G, Lejars O, Laithier V, Mechinaud F, Millot F, Kalifa C, Chastagner P. Topotecan as a radiosensitizer in the treatment of children with malignant diffuse brainstem gliomas: Results of a French Society of Paediatric Oncology Phase II Study. *Cancer*. 2005; 104:2792–2797. [PubMed: 16265674]
- Bryant J, Day R. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics*. 1995; 51:1372–1383. [PubMed: 8589229]
- Chen TT, Ng T. Optimal flexible designs in phase II clinical trials. *Statistics in Medicine*. 1998; 17:2301–2312. [PubMed: 9819829]
- Dent S, Zee B, Dancey J, Hanauske A, Wanders J, Eisenhauer E. Application of a new multinomial phase II stopping rule using response and early progression. *Journal of Clinical Oncology*. 2001; 19:785–791. [PubMed: 11157032]
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours; revised RECIST guideline (version 1.1). *European Journal of Cancer*. 2009; 45:228–247. [PubMed: 19097774]
- Escudier B, Grünwald V, Ravaud A, Ou YC, Castellano D, Lin CC, Gschwend JE, Harzstark A, Beall S, Pirota N, Squires M, Shi M, Angevin E. Phase II results of Dovitinib (TKI258) in patients with metastatic renal cell cancer. *Clinical Cancer Research*. 2014; 20:3012–3022. [PubMed: 24691021]
- Fleming TR. One-sample multiple testing procedures for phase II clinical trials. *Biometrics*. 1982; 38:143–151. [PubMed: 7082756]
- Glass B, Hasenkamp J, Wulf G, Dreger P, Pfreundschuh M, Gramatzki M, Silling G, Wilhelm C, Zeis M, Görlitz A, Pfeiffer S, Hilgers R, Truemper L, Schmitz N. Rituximab after lymphoma-directed conditioning and allogeneic stem-cell transplantation for relapsed and refractory aggressive non-Hodgkin lymphoma (DSHNHL R3): An open-label, randomised, phase 2 trial. *Lancet Oncology*. 2014; 15:757–766. [PubMed: 24827808]
- Gehan EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *Journal of Chronic Diseases*. 1961; 13:346–353. [PubMed: 13704181]
- Hainsworth JD, Infante JR, Spigel DR, Peyton JD, Thompson DS, Lane CM, Clark BL, Rubin MS, Trent DF, Burris HA 3rd. Bevacizumab and everolimus in the treatment of patients with metastatic melanoma: A phase 2 trial of the Sarah Cannon Oncology Research Consortium. *Cancer*. 2010; 116:4122–4129. [PubMed: 20564157]
- Ivanova A, Song G, Marchenko O, Moschos S. Monitoring rules for toxicity in phase II oncology trials. *Clinical Investigation*. 2015; 5(4):373–381.
- Li J, Juliar B, Yiannoutsos C, Ansari R, Fox E, Fisch MJ, Einhorn LH, Sweeney CJ. Weekly paclitaxel and gemcitabine in advanced transitional-cell carcinoma of the urothelium: A phase II Hoosier Oncology Group study. *Journal of Clinical Oncology*. 2005; 23:1185–1191. [PubMed: 15718315]
- London WB, Chang MN. One- and two-stage designs for stratified phase II clinical trials. *Statistics in Medicine*. 2005; 24:2597–2611. [PubMed: 16118809]
- Lu Y, Jin H, Lamborn KR. A design of phase II cancer trials using total and complete response endpoints. *Statistics in Medicine*. 2005; 24:3155–3170. [PubMed: 16189806]
- Mariani L, Marubini E. Content and quality of currently published phase II cancer trials. *Journal of Clinical Oncology*. 2000; 8:429–436.
- Miller DS, Blessing JA, Ramondetta LM, Pham HQ, Tewari KS, Landrum LM, Brown J, Mannel RS. Pemetrexed and cisplatin for the treatment of advanced, persistent, or recurrent carcinoma of the cervix: A limited access phase ii trial of the gynecologic oncology group. *Journal of Clinical Oncology*. 2014; 32:2744–2749. [PubMed: 25071133]

- Motzer RJ, Barrios CH, Kim TM, Falcon S, Cosgriff T, Harker WG, Srimuninnimit V, Pittman K, Sabbatini R, Rha SY, Flaig TW, Page R, Baybek S, Beck JT, Patel P, Cheung FY, Yadav S, Schiff EM, Wang X, Niolat J, Sellami D, Anak O, Knox JJ. Phase II randomized trial comparing sequential first-line everolimus and second-line sunitinib versus first-line sunitinib and second-line everolimus in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology*. 2014; 32:2765–2772. [PubMed: 25049330]
- Neuenschwander B, Rouyrre N, Hollaender N, Zuber E, Branson M. A proof of concept phase II non-inferiority criterion. *Statistics in Medicine*. 2011; 30:1618–1627. [PubMed: 21351286]
- Sargent DJ, Chan MD, Goldberg RM. A three-outcome design for phase II clinical trials. *Controlled Clinical Trials*. 2001; 22:117–125. [PubMed: 11306150]
- Schmoor C, Schumacher M, Finke J, Beyersmann J. Competing risks and multistate models. *Clinical Cancer Research*. 2013; 19:12–21. [PubMed: 23172882]
- Seymour L, Ivy SP, Sargent D, Spriggs D, Baker L, Rubinstein L, Ratain MJ, Le Blanc M, Stewart D, Crowley L, Groshen S, Humphrey JS, West P, Berry D. The design of phase II clinical trials testing cancer therapeutics: Consensus recommendations from the clinical trial design task force of the National Cancer Institute Investigational Drug Steering Committee. *Clinical Cancer Research*. 2010; 16:1764–1769. [PubMed: 20215557]
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians*. 2015; 65:5–29. [PubMed: 25559415]
- Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials*. 1989; 10:1–10. [PubMed: 2702835]
- Simon R, Wittes RE. Methodologic guidelines for reports of clinical trials. *Cancer Treatment Reports*. 1985; 69:1–3. [PubMed: 3967252]
- Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian NC, Gwyther SJ. New guidelines to evaluate the response to treatment in solid tumors (RECIST guidelines). *Journal of the National Cancer Institute*. 2000; 92:205–216. [PubMed: 10655437]
- Thezenas S, Duffour J, Culine S, Kramar A. Five-year change in statistical designs of phase II trials published in leading cancer journals. *European Journal of Cancer*. 2004; 40:1244–1249. [PubMed: 15110889]
- Vera-Badillo FE, Al-Mubarak M, Templeton AJ, Amir E. Benefit and harms of new anti-cancer drugs. *Current Oncology Reports*. 2013; 15:270–275. [PubMed: 23435854]

Table 1

Distributions of the 347 identified studies according to geographic characteristics.

	2005	2010	2014	Total
	(n = 141)	(n = 150)	(n = 56)	(n = 347)
Continent				
North America	83 (59%)	93 (62%)	30 (54%)	206 (59%)
Asia	9 (6%)	13 (9%)	4 (7%)	26 (7%)
Australia	3 (2%)	1 (1%)	1 (2%)	5 (1%)
Europe	46 (33%)	41 (27%)	19 (34%)	106 (31%)
Other	0 (0%)	2 (1%)	2 (4%)	4 (1%)
Study structure				
Multicenter	94 (67%)	110 (73%)	51 (91%)	255 (73%)
Single-Center	47 (33%)	40 (27%)	5 (9%)	92 (27%)
More than one country				
Yes	20 (14%)	31 (21%)	20 (36%)	71 (20%)
No	121 (86%)	119 (79%)	36 (64%)	276 (80%)
Cooperative oncology group				
US or Canada-based	25 (18%)	24 (16%)	8 (14%)	57 (16%)
European *	13 (9%)	3 (2%)	7 (13%)	23 (7%)
No	103 (73%)	123 (82%)	41 (73%)	267 (77%)

* Includes two studies from Australia.

Table 2

Distributions of the 347 identified studies according to types of cancer sites.

	<u>2005</u>	<u>2010</u>	<u>2014</u>	<u>Total</u>
	(n = 141)	(n = 150)	(n = 56)	(n = 347)
Lung				
SCLC	6 (4%)	6 (4%)	1 (2%)	13 (4%)
NSCLC	20 (14%)	18 (12%)	4 (7%)	42 (12%)
Gastrointestinal				
Colon	8 (6%)	9 (6%)	3 (5%)	20 (6%)
Pancreas	5 (4%)	2 (1%)	1 (2%)	8 (2%)
Esophagus	2 (1%)	2 (1%)	1 (2%)	5 (1%)
Liver	2 (1%)	1 (1%)	0 (0%)	3 (1%)
Other GI	2 (1%)	7 (5%)	2 (4%)	11 (3%)
Gynecological				
Breast	12 (9%)	22 (15%)	4 (7%)	38 (11%)
Ovary	4 (3%)	5 (3%)	3 (5%)	12 (3%)
Cervix	1 (1%)	1 (1%)	2 (4%)	4 (1%)
Genitourinary				
Bladder	2 (1%)	1 (1%)	4 (7%)	7 (2%)
Prostate	6 (4%)	7 (5%)	3 (5%)	16 (5%)
Testicular	1 (1%)	1 (1%)	2 (4%)	4 (1%)
Renal	2 (1%)	6 (4%)	3 (5%)	11 (3%)
Others				
Brain	5 (4%)	3 (2%)	0 (0%)	8 (2%)
Biliary tract	3 (2%)	3 (2%)	1 (2%)	7 (2%)
Head/neck	6 (4%)	5 (3%)	1 (2%)	12 (3%)
Glioblastoma	2 (1%)	2 (1%)	1 (2%)	5 (1%)
Thyroid	0 (0%)	4 (3%)	1 (2%)	5 (1%)
Leukemia	5 (4%)	2 (1%)	2 (4%)	9 (3%)
Lymphoma	10 (7%)	8 (5%)	8 (14%)	26 (7%)
Nervous system	0 (0%)	2 (1%)	0 (0%)	2 (1%)
Sarcoma	10 (7%)	3 (2%)	3 (5%)	16 (5%)
Skin	14 (10%)	13 (9%)	3 (5%)	30 (9%)
Other	13 (9%)	17 (11%)	3 (5%)	33 (10%)

SCLC, small-cell lung cancer; NSCLC, non-small-cell lung cancer; GI, gastrointestinal cancers.

Table 3

Primary Endpoints used in the 347 identified studies.

	2005	2010	2014	Total
	(n = 141)	(n = 150)	(n = 56)	(n = 347)
Response rate	83 (59%)	84 (56%)	24 (43%)	191 (55%)
CR, DC, CB	5 (4%)	16 (11%)	6 (11%)	27 (8%)
OS or PFS/TTP	31 (22%)	28 (19%)	23 (41%)	82 (24%)
Pathologic complete response *	4 (3%)	5 (3%)	1 (2%)	10 (3%)
PSA change **	4 (3%)	4 (3%)	0 (0%)	8 (2%)
Toxicity	2 (1%)	1 (1%)	1 (2%)	4 (1%)
Other	12 (9%)	12 (8%)	1 (2%)	25 (7%)

* Specific for neoadjuvant trials.

** Specific for prostate cancer trials.

CR, complete response; DC, disease control; CB, clinical benefit; OS, overall survival; PFS, progression-free survival; TTP, time to progression; PSA, prostatic specific antigen.

Table 4

Statistical design and number of patients per arm.

	2005	2010	2014	Total
	(n = 141)	(n = 150)	(n = 56)	(n = 347)
Multi-arm	50 (35%)	51 (34%)	28 (50%)	129 (37%)
Randomized				
Yes	31 (62%)	34 (67%)	23 (82%)	88 (25%)
No	19 (38%)	17 (33%)	5 (18%)	38 (11%)
Comparative				
Yes	20 (40%)	17 (33%)	9 (32%)	46 (36%)
No	30 (60%)	34 (67%)	19 (68%)	83 (64%)
Single arm	91 (65%)	99 (66%)	28 (50%)	218 (63%)
Statistical design				
Single stage	41 (45%)	44 (44%)	10 (36%)	95 (44%)
Simon	39 (43%)	35 (35%)	13 (46%)	87 (40%)
Fleming	2 (2%)	7 (7%)	3 (11%)	12 (6%)
Bayesian	1 (1%)	6 (6%)	0 (0%)	7 (3%)
Gehan	5 (6%)	1 (1%)	1 (3%)	7 (3%)
Flexible	0 (0%)	3 (3%)	0 (0%)	3 (1%)
Bryant & Day	2 (2%)	1 (1%)	0 (0%)	3 (1%)
Sargent	0 (0%)	2 (2%)	0 (0%)	2 (1%)
London	0 (0%)	0 (0%)	1 (4%)	1 (0%)
Triangular test	1 (1%)	0 (0%)	0 (0%)	1 (0%)
Number of patients per arm				
Mean (S.D.)	44.3 (24.7)	52.7 (38.6)	57.1 (39.8)	50.0 (34.1)
Median	39	45	45	41
Min–Max	10–187	10–269	16–235	10–269
<20	12 (9%)	13 (9%)	3 (5%)	28 (8%)
21–40	66 (47%)	52 (35%)	21 (38%)	139 (40%)
41–60	38 (27%)	53 (35%)	15 (27%)	106 (31%)
>60	25 (18%)	32 (21%)	17 (30%)	74 (21%)

Table 5

Hypothesis testing and study success.

	2005	2010	2014	Total
	(n = 141)	(n = 150)	(n = 56)	(n = 347)
Articles with statistical design “well described”				
Yes	119 (84%)	133 (89%)	49 (88%)	301 (87%)
No	22 (16%)	17 (11%)	7 (12%)	46 (87%)
Out of those with design “well described”, was the hypothesis formally tested?				
Yes	88 (74%)	108 (81%)	45 (93%)	241 (80%)
No	31 (26%)	25 (19%)	4 (7%)	60 (20%)
Out of those with formal hypothesis testing, was the hypothesis rejected/therapy recommended?				
Rejected and recommended	33 (38%)	50 (46%)	25 (56%)	109 (45%)
Not rejected but recommended	16 (18%)	27 (25%)	10 (22%)	53 (22%)
Not rejected and not recommended	38 (43%)	31 (29%)	10 (22%)	79 (33%)
Rejected but not recommended	1 (1%)	0 (0%)	0 (0%)	1 (0%)
Articles that recommended the therapy for further investigation among articles with statistical design “well described” and not “well described”				
Statistical design “well described”	72/119 (61%)	96/133 (72%)	38/49 (76%)	206/301 (68%)
Statistical design not “well described”	20/22 (91%)	16/17 (94%)	5/7 (71%)	41/46 (89%)

Table 6

Trials where statistical design was “well described”.

	N/Total (%)	p-Value
Overall	301/347 (87%)	
Year		
2005	119/141 (84%)	0.553
2010	133/150 (89%)	
2014	49/56 (88%)	
Continent		
North America	193/206 (94%)	<0.001
Europe	81/106 (76%)	
Other	27/35 (77%)	
Cooperative group		
Yes	74/80 (93%)	0.092
No	227/267 (85%)	
Patients per arm		
<40	141/167 (84%)	0.268
40	160/180 (89%)	