

# The Clinical Significance of Effect Sizes for Survival and Tumor Response Endpoints Using the Empirical Rule Effect Size

Major-Elechi BT<sup>1</sup>, Sloan JA<sup>1\*</sup>, Novotny PJ<sup>1</sup>, Sargent DJ<sup>1</sup>, Grothey A<sup>2</sup>, Lafky JM<sup>3</sup> and Dueck AC<sup>4</sup>

<sup>1</sup>Division of Biomedical Statistics and Bioinformatics, Mayo Clinic, Rochester, MN, USA

<sup>2</sup>Department of Medical Oncology, Mayo Clinic, Rochester, MN, USA

<sup>3</sup>Mayo Clinic, Rochester, MN, USA

<sup>4</sup>Department of Health Science Research, Mayo Clinic, Scottsdale Arizona, USA

## Abstract

**Context:** In planning oncology phase II and phase III clinical trials, the size of the expected effect for endpoints such as tumor response and overall survival are key parameters driving the sample size. We applied the empirical rule effect size approach, also known as the ½ standard deviation (SD) method, to define clinically significant effect sizes for overall survival and tumor response endpoints in a series of clinical trials.

**Methods:** The observed effect size was calculated for 12 phase II and 27 phase III completed cancer clinical trials identified by experts as being notable.

**Results:** The effect sizes of the phase II and phase III clinical trials ranged from -0.32 to 0.84 and 0.01 to 0.44 SDs respectively. Effect sizes for all but four of the phase II trials were less than a ½ SD. For phase III studies, the effect sizes for all but one study were below 0.4 SD and roughly 67% of them had an effect size smaller than 0.2 SDs. There were no differences across disease sites, although colorectal and breast trials did have slightly larger effect sizes.

**Conclusions:** Even highly noteworthy existing phase II and phase III oncology clinical trials rarely achieve the ½ SD level of clinical significance. This method allows for more ready interpretation of the clinical significance of overall survival and tumor response endpoints. It allows for cross-study comparison across different endpoints. The method also facilitates study design as it directly builds clinical significance into the study.

**Keywords:** Probability; Statistics in biology

## Introduction

Effect size is a key parameter for estimating sample size and power as well as interpreting results for clinical and observational studies [1]. An effect size should communicate information that is useful to assessing the clinical significance of any result found in a randomized clinical trial [2]. Traditional effect size calculation of expected overall survival (OS) times for phase III clinical trials involves preliminary estimates typically from previous trials [3]. Similarly, phase II clinical trial estimates of tumor response rates (RR) are usually derived from phase I clinical trials or pilot studies [4]. The clinical meaningfulness of the effect size used in phase II and III clinical trials is most often determined by investigator opinion rather than statistical calibration [5]. The statistical significance of a trial result, which is heavily dependent on sample size, size may vary from its inferred clinical significance. Principles to judge the clinical significance of a statistically significant effect size in different clinical contexts are necessary.

In order to provide a context for discussion summarizing OS and RR, we consider an area where clinical significance has been extensively studied; that of quality of life (QOL) data. Investigations involving QOL effect size assessments typically involve an additional hurdle in that the clinical significance of changes in QOL scores has been a subject of much debate [6]. Much has been written about various approaches for assessing clinical significance for QOL endpoints including anchor-based and distributional-based approaches [7].

Among the distribution based methods, one popular approach is the so-called ERES method (Empirical Rule Effect Size) or the ½ standard deviation method<sup>6</sup>. Based on a combination of Chebyshev's Theorem and Cohen's Effect size classifications, it has been demonstrated that in a broad spectrum of applications and techniques a between-groups difference of a ½ standard deviation is clinically non-ignorable [8,9]. A

smaller difference of a quarter or a third of a standard deviation may also have clinical meaning when other factors are considered; these smaller effect sizes are protocol-specific and should be defined a priori. Details of the mathematical underpinnings for the ERES method are in Sloan 2002. The ERES method allows for a consistent calibration of effect size across a wide spectrum across QOL assessments, treatment interventions and patient populations and allows for easy cross-study comparisons. The question arises as to whether this sort of cross-study comparison can be made for non-QOL endpoints such as overall survival and tumor response. This paper describes the development of such calibration and demonstrates its applicability in existing oncology clinical trials.

## Methods

### Mathematical underpinnings for the ½ standard deviation for overall survival

Assume the overall survival time of patient,  $x$ , follows an exponential distribution of  $f(x) = \frac{1}{t} e^{-\frac{x}{t}}$  where  $x \geq 0$  and  $t = \text{mean}$

**\*Corresponding author:** Sloan JA, Division of Biomedical Statistics and Bioinformatics, Mayo Clinic, Rochester, MN, USA, Tel: (507)284-9985; Fax: (507) 266-2477; E-mail: [averonik@cc.uoi.gr](mailto:averonik@cc.uoi.gr)

**Received** August 08, 2019; **Accepted** September 21, 2019; **Published** September 28, 2019

**Citation:** Major-Elechi BT, Sloan JA, Novotny PJ, Sargent DJ, Grothey A, et al. (2019) The Clinical Significance of Effect Sizes for Survival and Tumor Response Endpoints Using the Empirical Rule Effect Size. J Biom Biostat 10: 433.

**Copyright:** © 2019 Major-Elechi BT, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

overall survival time, then it follows directly that  $E(x)=t$ ,  $V_{ar}(x)=t^2$ ,  $SD(x)=t$  and finally that  $\frac{1}{2}SD(x)=\frac{t}{2}$ . Given that  $t$ =median overall survival time, then  $SD(x)=\frac{t}{\ln 2}$  and  $\frac{1}{2}SD(x)=\frac{t}{2\ln 2}$ .

### Mathematical underpinnings for the ½ standard deviation for tumor response

Assume the number successes,  $x$ , in a phase II clinical trial follows a binomial distribution with parameters  $n$  and  $p$  where  $p$  is probability of success. Then it follows directly that  $(x)=p$ ,  $V_{ar}(x)=p(1-p)$ ,  $SD(x)=\sqrt{p(1-p)}$  and finally that  $\frac{1}{2}SD(x)=\frac{1}{2}\sqrt{p(1-p)}$ . The sample proportion,  $\hat{p}=\frac{x}{n}$  can be used as an unbiased estimate of  $p$ .

### Mathematical underpinnings for the ½ standard deviation for hazard ratio

For phase III trials, the calibrated effect size is calculated as the difference in observed median overall survival between the two arms (i.e., observed treatment arm median  $OS[\hat{m}_t]$  minus observed reference arm median  $OS[\hat{m}_c]$  divided by the standard deviation of the reference arm median overall survival time computed as  $\frac{\hat{m}_c}{\ln 2}$ . Of note, the calibrated overall survival effect size  $\frac{(m_t - m_c)}{m_c / \ln 2}$  can be re-written as  $\ln 2 \left( \frac{m_t}{m_c} - 1 \right)$  showing that the effect size is also a function of the true inverse hazard ratio.

### Statistical analysis

The ERES calibration method can be applied to tumor response and overall survival endpoints. The calibrated effect size for overall survival is calculated as the difference in median overall survival divided by the standard deviation of the median overall survival time of the reference group. For the sake of emphasis in this paper, the observed difference in phase III clinical trials with three or more arms was taken as the difference between the reference group and the arm with the largest observed endpoint. In phase II trials the observed effect size was taken as the difference between the observed  $\hat{p}$  and the reference response rate taken from the power statement divided by the standard deviation of the reference response rate.

Sloan (2003) suggested that Cohen's effect sizes could be utilized as a means to define a non-ignorable clinically meaningful effect size. Small, medium, and large effect sizes can be characterized by the size of a worm so as to be insignificant (small), an elephant so as to be obvious to all (large), and a duck which from various perspectives could be recognized as clinically non-ignorable (medium). The duck allegory came from the quotation of Justice Warren Berger relating to the definition of pornography stating that it may be difficult to define but people know it when they see it. Similarly a clinically meaningful effect size should be non-ignorable from various viewpoints. Norman et al. furthered this work to show that a medium effect size or ½ SD is defensible from philosophical, clinical and even physiological perspectives.

For example, consider the following GlaxoSmithKlein phase III study (EGF104900; NCT00320385) where patients were randomized to receive single agent of lapatinib (1500 mg/daily) or a combination of lapatinib (1000 mg p.o. daily) plus trastuzumab (2 mg/kg). Women treated with monotherapy lapatinib experienced a median overall survival of 9.5 months compared with 14 months when treated with the

combination (median HR: 0.74,  $p=0.026$ ) [10]. This could be calibrated by saying that the effect is 0.36 standard deviations, which is a small/medium clinically meaningful effect size in this context although the  $p$ -value is statistically significant. The online only table gives further illustration for the ½ standard deviation calibration method for overall survival and tumor response endpoints.

To demonstrate the application of the ERES method, we reviewed phase II and phase III clinical trials conducted between the years 1999 and 2012 and identified thirty-nine exemplary clinical studies derived from the North Central Cancer Treatment Group (NCCTG), the Mayo Cancer Center, and a review conducted by Ocana and Tannock (2011) of clinical trials used by the United States Food and Drug Administration (FDA) for the approval of new molecular-targeted drugs. All studies reported median overall survival time or tumor response as a primary endpoint. The studies include the top three diagnosed cancers in the United States, as well as other cancers (7 breast cancer studies, 9 lung cancer studies, 12 gastrointestinal cancer studies and 11 other cancer studies). We further include material drawn from a recent review by the ASCO Cancer Research Committee for assessing clinically significant treatment effects<sup>11</sup>. Effect size was expressed as a multiple of the standard deviation (SD) with the following classifications: small effect size 0.2 SD, medium effect size 0.5 SD and large effect size 0.8 SD<sup>1</sup>.

### Results

Tables 1 and 2 present the key design parameters and observed results for the various phase II and phase III trials, respectively. For example in the first row of Table 1, results indicate for study NCCTG 954651, a phase II trial of oral 776C85 and oral 5-FU for patients with metastatic colorectal cancer, the study was designed with an assumed null response rate of 10% and an observed response rate of 26%. The calibrated effect size for this observed response rate is equivalent to 0.53 SD, which is slightly bigger than the non-ignorable ½ SD cut-off, which for this trial would have been 25% response rate. In Table 2, for example, the first row indicates that NCCTG 959255, a study of different agents for anorexia and cachexia, reported an observed difference of 18 days improvement in overall survival over the control group and a median survival of 120 days ( $p=0.66$ ). The calibrated effect size for this difference is 0.1 SD, a small effect size. In contrast, the WJLCG study by Furuse et al., comparing concurrent vs. sequential treatment for NSCLC, reported a 3.2 month improvement in survival ( $p=0.04$ ), which while statistically significant, represents a small calibrated effect size of 0.17 SD. Figure 1A presents effect sizes for the various phase II clinical trials making for a ready cross-study comparison of the magnitude of observed tumor response results. The effect sizes of the phase II clinical trials (Table 1) range from -0.32 SD (metastatic lung cancer) to 0.84 SD (metastatic breast cancer) and all but four trials have observed effects less than ½ standard deviation. A negative effect size indicates that the tumor response was less than the assumed null hypothesis. There were no substantial differences across disease sites.

Figure 1B presents effect sizes ranging from 0.01 SD to 0.44 SD for the phase III studies (Table 2), in order of magnitude. All effect sizes are below a ½ standard deviation threshold of a non-ignorable clinical effect with two thirds of the trials having effect sizes less than 0.2 SD. Phase III breast and colorectal cancer clinical trials had the largest effect sizes of 0.34 SD and 0.44 SD, respectively. Note, the effect sizes of all observed lung studies are less than 0.2 SD.

Ocana and Tannock performed a comprehensive review of phase III randomized clinical trials used by the FDA to approve new molecular-targeted drugs since 2000 [11,12].

Study no.	Description	Tumor sites	Study type	Reference response rate $p$	Observed response rate $\hat{p}$	Non-ignorable clinically significant response rate ( $\frac{1}{2}$ SD)	Calibrated observed effect size (SD)
954651 [16]	Phase II trial of oral 776C85 and oral 5-FU	GI	Metastatic	10%	26%	0.25	0.53
983252* [17]	Paclitaxel, carboplatin and trastuzumab for metastatic breast cancer	Breast	Adjuvant	40%	65%	0.64	0.51
983252* [17]	Paclitaxel, carboplatin and trastuzumab for metastatic breast cancer	Breast	Adjuvant	40%	81%	0.64	0.84
N0021* [18]	Gemcitabine and epirubicin for the treatment of mesothelioma	Head and Neck	Other: Locally Advanced	10%	13%	0.25	0.10
N0021* [18]	Gemcitabine and epirubicin for the treatment of mesothelioma	Head and Neck	Other: Locally Advanced	10%	7%	0.25	-0.10
N0022 [19]	Oral vinorelbine for the treatment of metastatic NSCLC	Lung	Metastatic	15%	3.4%	0.33	-0.32
N0044 [20]	Preoperative radiation and chemotherapy for locally advanced esophageal cancer	GI	Other: Neoadjuvant	40%	35%	0.64	-0.10
N0087* [21]	Interleukin-12 and rituximab in patients with non-Hodgkin's lymphoma	Lymphoma	Other	50%	37%	0.75	-0.26
N0087* [21]	Interleukin-12 in combination and rituximab in patients with non-Hodgkin's lymphoma	Lymphoma	Other	50%	52%	0.75	0.04
N0149 [22]	Oxaliplatin and Capecitabine for adenocarcinoma	GI	Cancer Control	15%	35%	0.33	0.56
N014C [23]	PS-341 and gemcitabine for pancreatic adenocarcinoma	Pancreas	Metastatic	5%	0%	0.16	-0.23
N0242 [24]	Docetaxel and capecitabine for adenocarcinoma	GI	Metastatic	20%	39%	0.40	0.48

\*Randomized phase II trial.

Table 1: Application of  $\frac{1}{2}$  standard deviation calibration method for tumor response to existing phase II NCCTG clinical trials included in the sample [16-24].

Study	Description	Tumor sites	Study type	Median overall survival of control group	Observed difference	Overall survival units	p-value	$\frac{1}{2}$ Standard deviation of control group	Calibrated effect size (SD)
NCCTG 959255 [25]	Megace vs. Marinol vs. both for anorexia and cachexia	GI, Lung, Other	Cancer Control	123	18	days	0.66	88.73	0.10
NCCTG 972451 [26]	CAI and placebo for advanced NSCLC	Lung	Metastatic	10.5	0.9	months	0.54	7.57	0.06
NCCTG 979251 [27]	Low molecular weight heparin for advanced cancer	Breast, GI, GU, Lung	Cancer Control	7.3	3.2	months	0.46	5.27	0.30
NCCTG N02C2 [28]	Erythropoietin in anemic patients with cancer	Various	Cancer Control	377	26	days	0.24	271.95	0.05
NCCTG N9841 [29]	CPT-11 versus OXAL/5-FU/CF for advanced colorectal carcinoma	GI	Metastatic	13.8	0.5	months	0.38	9.95	0.03
WJLCG [30]	Concurrent versus sequential treatment with radiotherapy (RT) and chemotherapy (CT) for NSCLC	Lung	Metastatic	13.3	3.2	months	0.04	9.59	0.17

RTOG 9410 [31]	Concurrent versus sequential treatment with radiotherapy (RT) and chemotherapy (CT) for NSCLC	Lung	Metastatic	14.6	2.4	months	Not reported	10.53	0.11
GMMA Ankara 1995 [32]	Concurrent versus sequential treatment with radiotherapy (RT) and chemotherapy (CT) for NSCLC	Lung	Metastatic	10	1	months	Not reported	7.21	0.07
GLOT-GFPC NPC 9501 [33]	Concurrent versus sequential treatment with radiotherapy (RT) and chemotherapy (CT) for NSCLC	Lung	Metastatic	14.5	1.8	months	0.24	10.46	0.09
EORTC 08972-22973 [34]	Concurrent chemo-radiotherapy versus sequential chemo-radiotherapy for inoperable NSCLC	Lung	Metastatic	16.2	0.3	months	Not reported	11.69	0.01
Irinotecan [35]	Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer	Colorectal/colon	Metastatic	4.3	2.7	months	0.004	3.1	0.44
Cetuximab NCT00079066 [36]	Cetuximab in patients with advanced colorectal cancer	Colorectal/colon	Metastatic	4.6	1.5	months	0.005	3.32	0.23
FOLFOX NCCTG N9741 [37]	Combinations of FU/LV, irinotecan and oxaliplatin in patients with previously untreated metastatic colorectal cancer	Colorectal/colon	Metastatic	15	4.5	months	0.0001	10.82	0.21
CLEOPATRA [38]	Combination of pertuzumab plus trastuzumab plus docetaxel as compared with placebo plus trastuzumab plus docetaxel when used as first-line treatment for HER2-positive metastatic breast cancer	Breast	Metastatic	12.4	6.1	months	<0.001	8.94	0.34
EMILIA [39]	T-DM1 compared with lapatinib plus capecitabine in patients with HER2-positive advanced breast cancer previously treated with trastuzumab and a taxane	Breast	Metastatic	25.1	5.8	months	<0.001	18.11	0.16

**Table 2:** Application of ½ standard deviation calibration method for overall survival to existing phase III clinical trials [25-39].

The ERES method was applied to the 12 studies that listed OS as a primary outcome (3 breast studies, 3 colorectal studies, 2 lung studies and 4 other cancer trials). Figure 2 and Table 3 demonstrate that all of these studies achieve less than a ½ standard deviation and 67% (8/12) of the studies had effect sizes <0.2 SD.

Recently, the American Society for Clinical Oncology (ASCO) Cancer Research Committee derived estimates for clinically meaningful differences in outcomes for cancer clinical trials based on subjective perception of ASCO members and expert reviewers [11]. Table 4 presents the results of that effort supplemented by estimates derived via our method. Column A presents what would be a ½ standard deviation effect size which has been suggested as a non-ignorable effect [13]. The ½ SD effect sizes in Table 4 are all substantially larger than

the upper estimates arrived at by expert review, with the exception of colon cancer. Hence the size of treatment effect that clinicians would consider as clinically meaningful for most cancer clinical trials is much smaller than ½ SD. Column B of Table 4 indicates that in pancreatic cancer a calibrated effect size of between 0.25 and 0.46 SD and in colon cancer a calibrated effect size between 0.35-0.87 SD could be considered clinically significant, compared to effect sizes of around 0.20 SD, for breast and lung cancer.

## Discussion

The present article provides a calibration method for interpreting results of clinical trials within the context of similar trials in the same disease and even across disease sites. This ERES calibration method

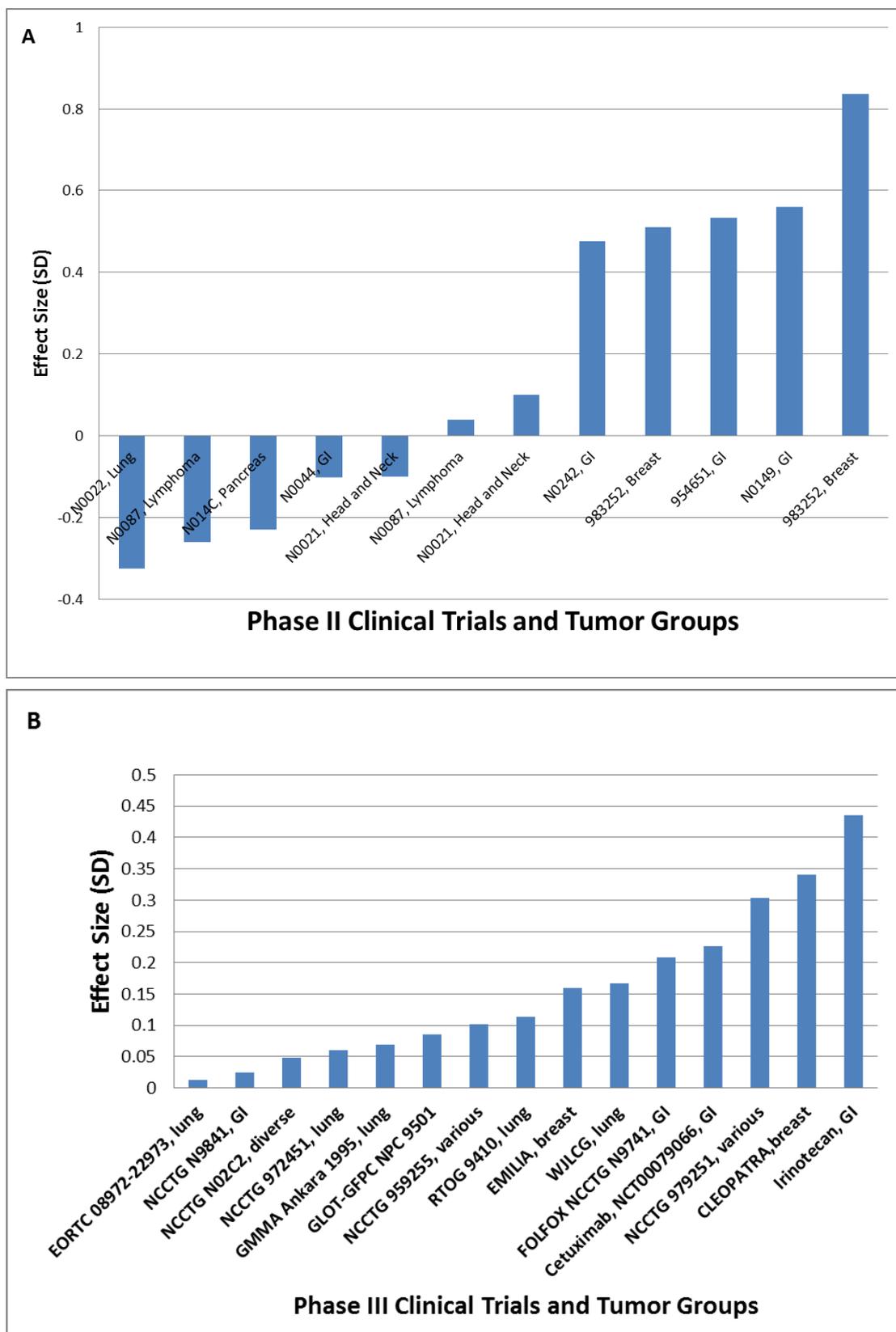
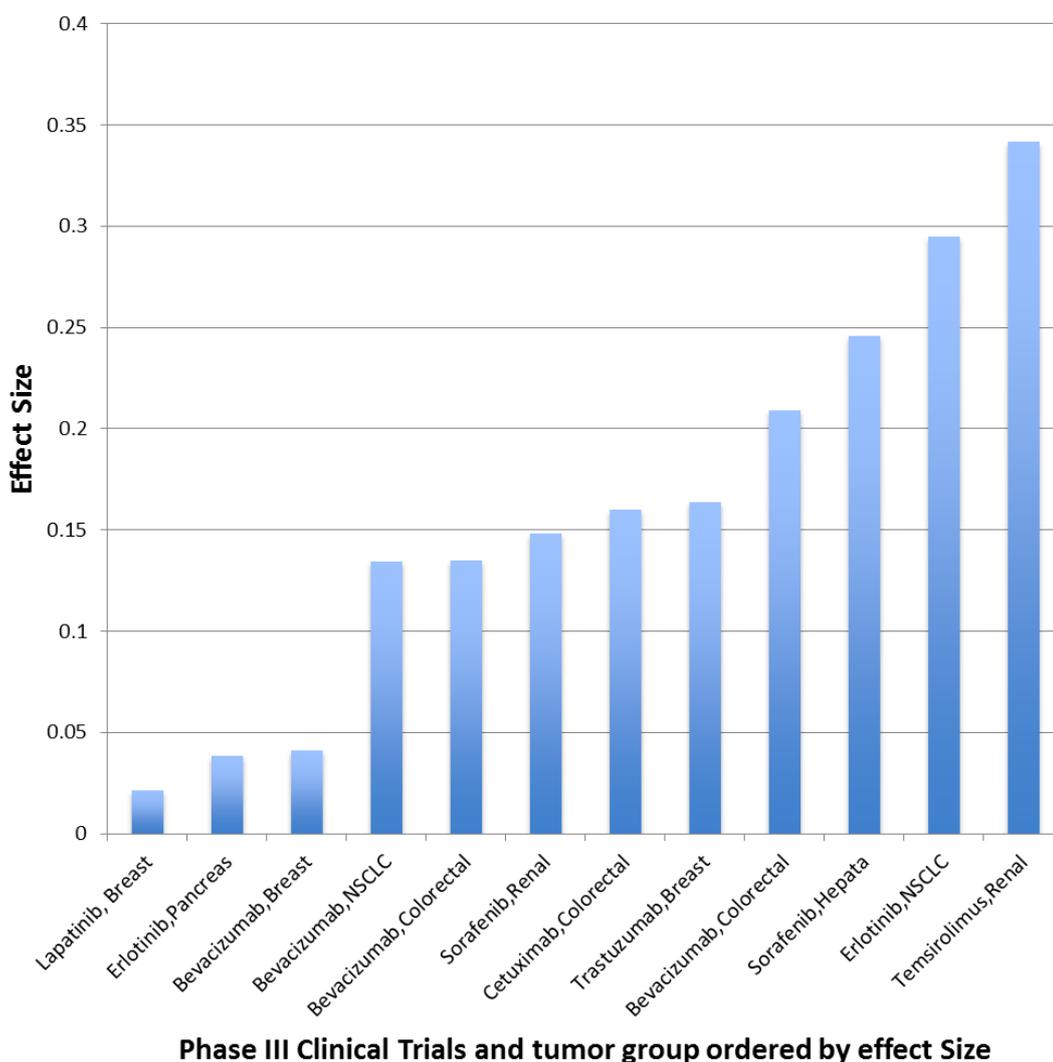


Figure 1: Calibrated effect sizes for tumor response in phase II (A) and median OS in phase III (B) NCCTG clinical trials ordered by size of effect.



**Figure 2:** Calibrated effect sizes for overall survival of phase III clinical trials used by the FDA for the approval of new molecular-targeted drugs since 2000 ordered by size of effect (Ocana and Tannock [12]).

Author, Year	Drug	Tumor site	Median overall survival (months) of control group	Observed difference (months)	p-value	½ Standard deviation of control group (months)	Calibrated effect size (SD)
Johnston, 2009	Lapatinib	Breast	32.3	1	0.019	23.3	0.02
Slamon, 2001	Trastuzumab	Breast	20.3	4.8	0.001	14.6	0.16
Miller, 2007	Bevacizumab	Breast	25.2	1.5	0.001	18.2	0.04
Hurwitz, 2004	Bevacizumab	Colorectal	15.6	4.7	0.001	11.3	0.21
Giantonio, 2007	Bevacizumab	Colorectal	10.8	2.1	0.001	7.8	0.13
Jonker, 2007 [36]	Cetuximab	Colorectal	6.5	1.5	0.001	4.7	0.16
Moore, 2007	Erlotinib	Pancreas	5.91	0.33	0.038	4.3	0.04
Llovet 2008	Sorafenib	Hepatocellular	7.9	2.8	0.001	5.7	0.25
Shepherd, 2005	Erlotinib	NSCLC	4.7	2	0.001	3.4	0.29
Sandler, 2006	Bevacizumab	NSCLC	10.3	2	0.003	7.4	0.13
Escudier, 2007	Sorafenib	Renal	15.9	3.4	0.02	11.5	0.15
Hudes, 2007	Temsirolimus	Renal	7.3	3.6	0.008	5.3	0.34

**Table 3:** Characteristics and calibrated effect sizes for overall survival of phase III randomized clinical trials used by the United States Food and Drug Administration (FDA) for the approval of new molecular-targeted drugs during 2000-2010 [12].

Cancer type	Patient population	Current baseline median overall survival	Improvement over current overall survival that would be clinically meaningful*	½ Standard deviation (column A)	Calibrated effect size (SD) (column B)
Pancreatic Cancer	FOLFIRONOX Eligible Patients	10-11 months	4-5 months	7.21-7.93 months	0.25-0.35
Pancreatic Cancer	Gemcitabine Eligible Patients	6-8 months	3-4 months	4.33-5.77 months	0.26-0.46
Lung Cancer	Non-squamous cell carcinoma	13 months	3.25-4 months	9.38 months	0.17-0.21
Lung Cancer	Squamous cell carcinoma	10 months	2.5-3 months	7.21 months	0.17-0.21
Breast Cancer	Metastatic triple negative, previously untreated for metastatic disease	18 months	4.5-6 months	12.98 months	0.17-0.23
Colon Cancer	Disease progression on all prior therapies (or not a candidate for standard 2 <sup>nd</sup> or 3 <sup>rd</sup> line options)	4-6 months	3-5 months	2.89-4.33 months	0.35-0.87

\*American Society for Clinical Oncology (ASCO) Cancer Research Committee derived estimates [11].

**Table 4:** Summary of recommended overall survival targets for meaningful clinical trial goals.

has previously been used to benchmark QOL outcomes. However, the empirical data reflecting the results of recent oncology clinical trials analyzed here indicated that under the vast majority of circumstances, the calibrated effect sizes of these oncology clinical trials are less than ½ standard deviation. The ½ standard deviation criterion is not an arbitrary value but has been empirically derived based on a combination of Chebyshev's Theorem and the pervasive "Cohen's d" effect size. However, a limitation of the method is that it is dependent on the distribution of the survival function and also the proportional hazards assumption.

It has been previously demonstrated that the effect sizes of recent 'positive' randomized controlled clinical trials are smaller than those in the past and that the main predictor of a positive clinical trial is a statistically significant result [14]. Although 19 of the 27 phase III trials were statistically significant, none had effect sizes >0.5 SD and almost all (25 of 27) had effect sizes <0.2 SD. Four of the 12 phase II trials had effect sizes >0.5 SD. It is therefore imperative to develop principles that underlie the clinical significance threshold in different clinical contexts. It also opens up the question as to whether a ½ SD rule is practical for oncology trials with an OS endpoint, since it may not be achievable for survival outcomes. This is not to discourage or disparage recent and future oncology clinical trials, but hopefully put expectations into perspective. Further, the method still allows for comparative calibration across clinical trials involving survival outcomes.

A further perspective that is challenged by this work is the extensive discussion and conclusion regarding effect sizes for QOL outcomes. Often it has been noted that the size of effects seen in QOL studies is less than a ½ standard deviation and is declared a clinically insignificant result. However, if the same rubric were to be applied to tumor response and overall survival effect sizes, the vast majority of those results would also be characterized as clinically non-significant. This opens the discussion to a potential re-examination of how one can interpret clinical trial results for overall survival and tumor response outcomes versus QOL-related domains and toxicity. The ERES method can be helpful here, as described elsewhere, in the development of a quality adjusted survival metric which combines survival and toxicity/QOL into a single summary statistic [15].

## Conclusion and Implications

The ERES method allows for a simple and mathematically consistent interpretation of the clinical significance of overall survival and tumor response studies. It allows for direct cross-study comparison across

different end points and facilitates study design as it builds clinical significance into the study directly. The primary advantage of the ERES method is that it allows for a calibration of design parameters and trial results into a common metric. For example, when comparing results of two clinical trials, one might encounter an improvement of three months of overall survival that represents a small effect size in one trial and a moderate effect size in another. This is consistent with the findings of the ASCO committee that an additional month of overall survival could be a large improvement (as in metastatic pancreatic cancer) or a small improvement (as in early stage breast disease).

The next step in this line of research (in progress) is to produce real time applications that can facilitate effective communication among clinical trial development teams.

Ready-made applications for electronic devices could be employed to translate previous literature and subjective estimates of efficacy into data for constructing and comparing alternative statistical design parameters. Ultimately, the success of this approach will lie in the ability for the abstract effect size approach to be understood and readily used by individuals involved in the design and interpretation of clinical trials.

## Acknowledgements

Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10CA180821, U10CA180882, and UG1CA189823 to the Alliance for Clinical Trials in Oncology; CA25224 to the legacy North Central Cancer Treatment Group (NCCTG); and CA-37404, CA-35431, CA-35415, CA-35103, and CA-35269. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

- Cohen J (1988) *Statistical Power analysis for the behavioral sciences*. Psychology Press: Hillsdale.
- US Department of Health and Human Services FaDA (1998) Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. Guidance for Industry E9 Statistical Principles for Clinical Trials.
- Kraemer HC, Mintz J, Noda A (2006) Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry* 63: 484-489.
- Cannistra SA (2009) Phase II trials in journal of clinical oncology. *J Clin Oncol* 27: 3073-3076.
- Altman DG (1990) *Practical Statistics for Medical Research*. Chapman and Hall/CRC
- Sloan JA, Dueck A (2004) Issues for statisticians in conducting analyses and translating results for quality of life end points in clinical trials. *J Biopharm Stat* 14: 73-96.

7. Revicki DA, Erickson PA, Sloan A, Dueck A, Guess H, et al. (2007) Interpreting and reporting results based on patient-reported outcomes. *Value Health* 10 Suppl 2: S116-124.
8. Norman GR, Sloan JA, Wyrwich KW (2003) Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Med Care* 41: 582-592.
9. Sloan JA, Cella D, Hays RD (2005) Clinical significance of patient-reported questionnaire data: another step toward consensus. *J Clin Epidemiol* 58: 1217-1219.
10. Blackwell KL, Burstein HJ, Storniolo AM, Rugo H, Sledge G, et al. (2012) Overall survival benefit with lapatinib in combination with trastuzumab for patients with human epidermal growth factor receptor 2-positive metastatic breast cancer: final results from the EGF104900 Study. *J Clin Oncol* 30: 2585-2592.
11. Ellis LM, Bernstein DS, Voest EE, Berlin JD, Sargent D, et al. (2014) American Society of Clinical Oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. *J Clin Oncol* 32: 1277-1280.
12. Ocana A, Tannock IF (2011) When are "positive" clinical trials in oncology truly positive? *J Natl Cancer Inst* 103: 16-20.
13. Sloan JA, Zhao X, Novotny PJ, Wampfler J, Garces Y, et al. (2012) Relationship between deficits in overall quality of life and non-small-cell lung cancer survival. *J Clin Oncol* 30: 1498-1504.
14. Booth CM, Cescon DW, Wang L, Tannock IF, Krzyzanowska MK (2008) Evolution of the randomized controlled trial in oncology over three decades. *J Clin Oncol* 26: 5458-5464.
15. Major-Elechi BT, Novotny PJ, Singh JA, Bonner JA, Dueck AC, et al. (2008) Combining Survival and Toxicity Effect Sizes from Clinical Trials: NCCTG 89-20-52 (Alliance). *Int J Stat Med Res* 7: 137-146.
16. Hobday TJ, Kugler JW, Mahoney MR, Sargent DJ, Sloan JA, et al. (2002) Efficacy and quality-of-life data are related in a phase II trial of oral chemotherapy in previously untreated patients with metastatic colorectal carcinoma. *J Clin Oncol* 20: 4574-4580.
17. Perez EA, Suman VJ, Rowland KM, Ingle JN, Salim M, et al. (2005) Two concurrent phase II trials of paclitaxel/carboplatin/trastuzumab (weekly or every-3-week schedule) as first-line therapy in women with HER2-overexpressing metastatic breast cancer: NCCTG study 983252. *Clin Breast Cancer* 6: 425-432.
18. Okuno SH, Delaune R, Sloan JA, Foster NR, Maurer MJ, et al. (2008) A phase 2 study of gemcitabine and epirubicin for the treatment of pleural mesothelioma: a North Central Cancer Treatment Study, N0021. *Cancer* 112: 1772-1779.
19. Kanard A, Jatoi A, Castillo R, Geyera S, KSchulz T, et al. (2004) Oral vinorelbine for the treatment of metastatic non-small cell lung cancer in elderly patients: a phase II trial of efficacy and toxicity. *Lung Cancer* 43: 345-353.
20. Jatoi A, Martenson JA, Foster NR, McLeod HL, Lair BS, et al. (2007) Paclitaxel, carboplatin, 5-fluorouracil, and radiation for locally advanced esophageal cancer: phase II results of preliminary pharmacologic and molecular efforts to mitigate toxicity and predict outcomes: North Central Cancer Treatment Group (N0044). *Am J Clin Oncol* 30: 507-513.
21. Ansell SM, Geyer SM, Maurer MJ, Kurtin PJ, Micallef INM, et al. (2006) Randomized phase II study of interleukin-12 in combination with rituximab in previously treated non-Hodgkin's lymphoma patients. *Clin Cancer Res* 12: 6056-6063.
22. Jatoi A, Murphy BR, Foster NR, Nikcevich DA, Alberts SR, et al. (2006) Oxaliplatin and capecitabine in patients with metastatic adenocarcinoma of the esophagus, gastroesophageal junction and gastric cardia: a phase II study from the North Central Cancer Treatment Group. *Ann Oncol* 17: 29-34.
23. Alberts SR, Foster NR, Morton RF (2005) PS-341 and gemcitabine in patients with metastatic pancreatic adenocarcinoma: a North Central Cancer Treatment Group (NCCTG) randomized phase II study. *Ann Oncol* 16: 1654-1661.
24. Giordano KF, Jatoi A, Stella PJ, Foster N, Tschetter LK, et al. (2006) Docetaxel and capecitabine in patients with metastatic adenocarcinoma of the stomach and gastroesophageal junction: A phase II study from the North Central Cancer Treatment Group. *Ann Oncol* 17: 652-656.
25. Jatoi A, Windschitl HE, Loprinzi CL, Sloan JA, Dakhil SR, et al. (2002) Dronabinol versus megestrol acetate versus combination therapy for cancer-associated anorexia: A North Central Cancer Treatment Group study. *J Clin Oncol* 20: 567-573.
26. Johnson EA, Marks RS, Mandrekar SJ, Hillman SL, Haugec MD, et al. (2008) Phase III randomized, double-blind study of maintenance CAI or placebo in patients with advanced non-small cell lung cancer (NSCLC) after completion of initial therapy (NCCTG 97-24-51). *Lung Cancer* 60: 200-207.
27. Sideras K, Schaefer PL, Okuno SH, Sloan JA, Kutteh L, et al. (2006) Low-molecular-weight heparin in patients with advanced cancer: A phase 3 clinical trial. *Mayo Clinic proceedings* 81: 758-767.
28. Steensma DP, Molina R, Sloan JA, Nikcevich DA, Schaefer PL, et al. (2006) Phase III study of two different dosing schedules of erythropoietin in anemic patients with cancer. *J Clin Oncol* 24: 1079-1089.
29. Kim GP, Sargent DJ, Mahoney MR, Rowland KM, Philip PA, et al. (2009) Phase III noninferiority trial comparing irinotecan with oxaliplatin, fluorouracil, and leucovorin in patients with advanced colorectal carcinoma previously treated with fluorouracil: N9841. *J Clin Oncol* 27: 2848-2854.
30. Furuse K, Fukuoka M, Kawahara M, Nishikawa H, Takada Y, et al. (1999) Phase III study of concurrent versus sequential thoracic radiotherapy in combination with mitomycin, vindesine, and cisplatin in unresectable stage III non-small-cell lung cancer. *J Clin Oncol* 17: 2692-2699.
31. Curran WJ (2003) Long-term benefit is observed in a phase III comparison of sequential versus concurrent chemo-radiation for patients with unresected stage III NSCLC: RTOG 94-10.
32. Ulutin HC, Guden M, Oysul K, Surenkoc S, Pak Y, et al. (2000) Split-course radiotherapy with or without concurrent or sequential chemotherapy in non-small cell lung cancer. *Radiat Med* 18: 93-96.
33. Fournel P, Robinet G, Thomas P, Souquet PJ, Léna H, et al. (2005) Randomized phase III trial of sequential chemoradiotherapy compared with concurrent chemoradiotherapy in locally advanced non-small-cell lung cancer: Groupe Lyon-Saint-Etienne d'Oncologie Thoracique-Groupe Français de Pneumo-Cancerologie NPC 95-01 Study. *J Clin Oncol* 23: 5910-5917.
34. Belderbos J, Uitterhoeve L, van Zandwijk N, Belderbos H, Rodrigus P, et al. (2007) Randomised trial of sequential versus concurrent chemo-radiotherapy in patients with inoperable non-small cell lung cancer (EORTC 08972-22973). *Eur J Cancer* 43: 114-121.
35. Saltz LB, Cox JV, Blanke C, Rosen LS, Fehrenbacher L, et al. (2000) Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *N Engl J Med* 343: 905-914.
36. Jonker DJ, O'Callaghan CJ, Karapetis CS, Zalberg JR, Tu D, et al. (2007) Cetuximab for the treatment of colorectal cancer. *N Engl J Med* 357: 2040-2048.
37. Goldberg RM, Sargent DJ, Morton RF, Fuchs CS, Ramesh K, et al. (2006) Randomized controlled trial of reduced-dose bolus fluorouracil plus leucovorin and irinotecan or infused fluorouracil plus leucovorin and oxaliplatin in patients with previously untreated metastatic colorectal cancer: a North American Intergroup Trial. *J Clin Oncol* 24: 3347-3353.
38. Baselga J, Cortes J, Kim SB, Seock-Ah Im, Hegg R, et al. (2012) Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 366: 109-119.
39. Verma S, Miles D, Gianni L, Krop IE, Welslau M, et al. (2012) Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med* 367: 1783-1791.