ORIGINAL RESEARCH

Number Of Clinical Trial Study Sites Impacts Observed Treatment Effect Size: An Analysis Of Randomized Controlled Trials Of Opioids For Chronic Pain

Diana S Meske [1]
Ben J Vaughn [2]
Ernest A Kopecky [3]
Nathaniel Katz [4,5]

[1]Collegium Pharmaceutical, Inc, Stoughton, MA, USA; [2]Rho, Chapel Hill, NC, USA; [3]Teva Branded Pharmaceutical Products R&D, Inc, Frazer, PA, USA; [4]WCG Analgesic Solutions, Wayland, MA, USA; [5]Department of Anesthesia, Tufts University School of Medicine, Boston, MA, USA

Background: Many aspects of study conduct impact the observed effect size of treatment. Data were utilized from a recently published meta-analysis of randomized, double-blind, placebo-controlled, clinical trials performed for the United States Food and Drug Administration (FDA) approval of full mu-agonist opioids for the treatment of chronic pain.

Methods: The number of study sites in each clinical trial and standardized effect size (SES) were extracted and computed. Standardized effect size was plotted against number of sites, and a two-piece linear model was fit to the plot. Ten studies were included.

Results: The SES decreased linearly by 0.13 units for every 10 sites (p=0.037), from 0.75 to 0.36, until an inflection point of 60 sites, after which SES did not decline further. The total number of subjects required for 90% power to discriminate drug from placebo increased from 78 to 336 subjects going from 30 to 60 sites.

Conclusion: Results showed that the number of sites was a source of loss of assay sensitivity in clinical trials, which may contribute to the well-known problem of failure to successfully transition from Phase 2 to Phase 3 clinical development. Potential solutions include minimizing the number of sites, more rigorous and validated training, central statistical monitoring with rapid correction of performance issues, and more rigorous subject and site selection.

Keywords: randomized controlled trials, opioids, chronic pain, clinical trials, effect size, study site number

Introduction

The United States Food and Drug Administration (FDA) requires randomized, double-blind, placebo-controlled studies of at least a 3-month duration to approve treatments for chronic diseases such as pain.[1] These studies typically take years to enroll and pose a significant financial burden on companies and institutions seeking regulatory approval. To expedite enrollment, many sponsors are tempted to utilize large numbers of study sites. However, inconsistencies in study conduct at each study site have long been assumed to add variability to the data[2] and may increase the likelihood of a failed study (i.e., the treatment actually is efficacious, but the study fails).[3]

Results of small, carefully performed, successful Phase 2 studies have long been observed to be difficult to replicate in large, multi-site, Phase 3 studies, even when

Correspondence: Diana S Meske
Collegium Pharmaceutical, Inc, Stoughton, MA, USA
Tel +1 517 712 3087
Email dsmeske@gmail.com

sample size is increased.[4] There are two possible explanations: 1) the smaller Phase 2 study results were falsely positive or 2) due to larger error variability, the Phase 3 study was falsely negative, or lacked adequate assay sensitivity. Opioids are a useful class of drugs for discerning the impact of study design and conduct on observed effect sizes because, in principle, all full mu-opioid agonists have the same pharmacologic mechanism, binding to the mu-opioid receptor, and therefore produce the same therapeutic effect once doses are optimized.[5] A 2005 qualitative review of all randomized controlled trials (RCTs) of opioids found that the fewer the sites, the larger the observed standardized effect size (SES) of treatment.[6] The relationship between numbers of sites and observed treatment effects has seldom been quantitatively evaluated.

This study reanalyzes data previously collected for a meta-analysis of clinical trials of opioids versus placebo for chronic pain submitted for regulatory approval in the US,[7] to characterize the relationship between number of sites and SES, and to suggest an appropriate adjustment to SES for sample size calculations based on the number of study sites the sponsor intends to use in a Phase 3 clinical trial.

Materials And Methods

This analysis utilized data collected from a recent meta-analysis of opioid studies for chronic pain.[7] A detailed description of the methods of the original study can be found in the original publication.

In this study, all published double-blind, randomized, placebo-controlled, enriched enrollment randomized withdrawal (EERW) studies that compared any full mu-agonist opioid (excluding partial mu-agonists and mixed activity opioids) to placebo for ≥12 weeks during the randomized, double-blinded treatment phase were selected. Any type of chronic non-cancer pain was allowed; acute or cancer pain was excluded. Agents with oral, transdermal, nasal, sublingual, and transmucosal routes of administration were included.

The original study protocol was published on PROSPERO (registration #: CRD42015026378).

The primary endpoint of the current analysis was the relationship across studies between the number of study sites and SES. The number of study sites in which each clinical trial was conducted was extracted, as were sufficient data to compute an SES (change in pain intensity [PI] from randomization baseline to week 12 or study endpoint and its variability); when necessary, this information was extracted from other publicly available sources (e.g., FDA Summary Basis of Approval, www.clinicaltrials.gov). Standardized effect size for each study was defined as [{(change from baseline to week 12 in PI on drug) – (change from baseline to week 12 in PI on placebo)}/ (pooled standard deviation of the change from baseline in PI)]. SES is independent of the study's sample size, but studies with larger sample sizes will have more precise estimates of the SES.

Observed SES was plotted against the number of sites utilized for each protocol and a smoothed line was fit to the plot to examine the overall shape. For sensitivity, this was jackknifed by removing one protocol at a time from the sample and repeating the analysis. This revealed a consistent inflection point at around 60 sites followed by a flat trend line beyond 60 sites. This was then used for selection of the knot point in the final two-piece linear model. The number of subjects that theoretically would be required for 90% power to detect the observed SES at participating sites was then calculated based on the model-predicted SES for a given site count and plotted.

Results

Fifteen EERW studies were included in the original meta-analysis. Five were excluded because they studied drugs were not full mu-agonists or had additional mechanisms of action (monoamine reuptake inhibition), namely buprenorphine, tramadol, and tapentadol, leaving 10 studies (see Table 1). Drugs evaluated were hydrocodone (4 clinical trials); oxymorphone and oxycodone (2 clinical trials each); and hydromorphone and morphine/naltrexone (1 clinical trial each). Disorders studied included chronic low back pain and osteoarthritis. Standardized effect sizes were based on the group mean difference between drug and placebo at the end of the randomized, double-blind treatment period, and ranged from 0.173 to 0.913. The estimated mean SES of the 10 studies included in this analysis as assessed by binary random-effects model meta-analysis using the restricted maximum likelihood method is 0.399 (95% CI: 0.239, 0.559). Further details are available in the original publication.[7]

The relationship between number of sites and observed SES is presented graphically in Figure 1, which also illustrates the results of the two-piece linear model (p-value=0.037). The slope of the model declined at a rate of 0.13 SES units per 10 sites from approximately 30 to 60 sites (i.e., for every additional 10 sites, the SES declined by 0.13 units), then was flat (i.e., adding

Table I Enriched Enrollment, Randomized Withdrawal Studies Of Full Mu Opioid Agonists For The Treatment Of Chronic Pain Submitted For FDA Approval Included In This Analysis

Author Year	Drug, Pain Type	N	# Of Sites, Countries	Change In PI-NRS Score Mean (SD)		SES Estimate (95% CI)
Hale[8] 2007	Oxymorphone ER, CLBP	142	30, US	Drug: Placebo:	8.7 (25.1) 31.6 (24.6)	0.913 (0.567, 1.259)
Katz[9] 2007	Oxymorphone ER, CLBP	205	29, US	Drug: Placebo:	10.9 (24.5) 26.0 (27.9)	0.574 (0.294, 0.853)
Hale[10] 2010	Hydromorphone ER, CLBP	266	66, US	Drug: Placebo:	0.6 (1.3) 1.7 (1.5)	0.799 (0.550, 1.049)
Katz[11] 2010	Morphine – sequestered naltrexone ER capsules, OA	343	74 US	Drug: Placebo:	−0.2 (1.9) 0.3 (2.1)	0.249 (0.037, 0.462)
Friedman[12] 2011	Oxycodone ER, OA	410	61, US	Drug: Placebo:	−0.7 (2.1) −0.3 (2.5)	0.173 (−0.021, 0.367)
Rauck[13] 2014	Hydrocodone ER, CLBP	302	59, US	Drug: Placebo:	0.48 (1.6) 0.96 (1.6)	0.308 (0.081, 0.535)
Wen[14] 2015	Hydrocodone bitartrate ER, CLBP	588	102, US	Drug: Placebo:	3.7 (1.9) 4.2 (2.0)	0.267 (0.104, 0.429)
Katz[15] 2015	Oxycodone ER, CLBP	389	46, US	Drug: Placebo:	0.29 (2.1) 1.85 (3.1)	0.592 (0.389, 0.795)
Hale[16] 2015a	Hydrocodone ER, CLBP/OA	293	71, US	Drug: Placebo:	−0.64 (1.6) −0.03 (1.9)	0.347 (0.117, 0.578)
Hale[17] 2015b	Hydrocodone ER, CLBP	371	78, US	Drug: Placebo:	0.11 (1.9) 0.74 (2.0)	0.319 (0.114, 0.524)

Note: Data from Meske et al.[7]
Abbreviations: CLBP, chronic low back pain; CI, confidence interval; ER, extended-release; N, number of subjects included in the randomized, double-blind, placebo-controlled portion of the study; OA, osteoarthritis; PI-NRS, pain intensity – numerical rating scale; SD, standard deviation; SES, standardized effect size; US, United States.

additional sites was not associated with a further decline in SES). Because the smallest number of sites of any of the included studies was 29, there are no data below that number of sites.

Figure 2 presents the total number of subjects required for 90% power to detect the observed difference under the model shown in Figure 1, contrasted with the number needed under the traditional assumption of no impact of number of sites on observed effect size. As clinical trial size increases from 30 to 60 sites, the total number of subjects required increases from 78 to 336. Beyond 60 sites, there is no further increase in sample size requirement because there was no further degradation of SES.
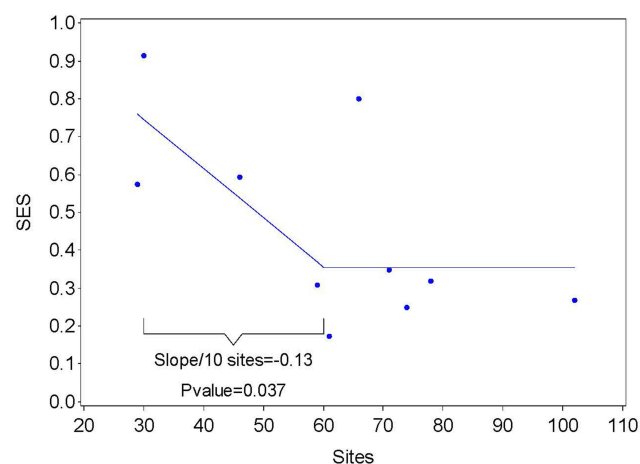

Figure I A two-piece linear model of the relationship between number of study sites (x-axis) and observed SES of treatment (y-axis).

Discussion

The problem of variability in study conduct between clinical trial sites has been lamented for years in both scientific[18–21] and regulatory[2,22] documents. There appears to be only one previous observation on the relationship between number of sites and effect sizes, a meta-analysis
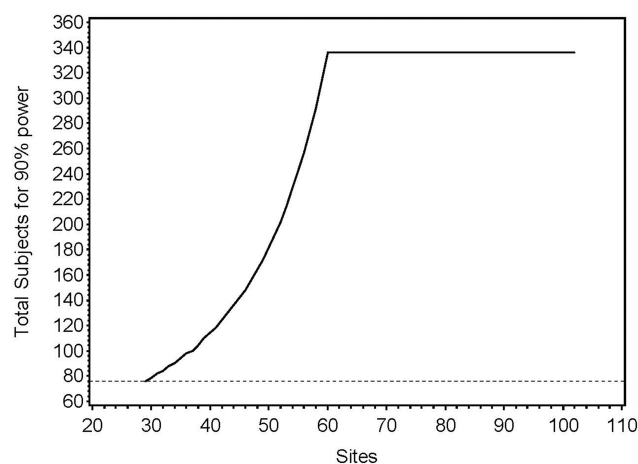
Figure 2 Total number of subjects required form 90% power, comparing the modeled actual relationship between number of sites and SES, vs the assumption of no impact of number of sites in SES.

of pediatric antidepressant clinical trials, which mentioned that the effect size was inversely proportional to the number of sites.[23] To the authors' knowledge, the current study is the first study to focus on the quantitative relationship between number of sites and effect sizes. Results showed that as the number of sites in clinical trials of opioids for chronic pain increases, the effect size decreases, until a plateau is reached where effect size does not decline further. The impact was large: a 30-site study, fully powered with 80 subjects, would need over 300 subjects if the number of sites were increased to 60. Thus, using data from Phase 2 studies conducted with fewer sites may seriously underestimate sample size requirements when applied to Phase 3 clinical trials with more sites.

While there are good reasons to conduct multicenter clinical trials,[19] utilizing many sites to accelerate timelines may lead to a "haste makes waste" phenomenon, where by racing to complete clinical trials as soon as possible, sponsors may end up at the starting line, as failed clinical trials need to be repeated. The ethical implications of experimenting on human subjects in inadvertently underpowered studies are also concerning. Moreover, some sponsors may not be able to withstand the financial consequences of a failed study; promising treatments may never make it to market.

The measurement error that underlies the impact of site variability on assay sensitivity must be minimized to avoid false-negative clinical trials.[3] Several studies have described methods to prevent measurement error by training patients to report their symptoms more accurately,[24] screening out patients who are not able to report symptoms

accurately,[25,26] and a variety of other methods.[3] Identification and remediation of measurement error during ongoing clinical trials by central statistical monitoring and intervention, as required by regulation,[22] has also been described in a few studies.[18,27]

This analysis was limited by the number of available studies; the observed relationship could change with more studies. Only opioid clinical trials were included; clinical trials of other analgesics, or other therapeutic areas, may not show the same relationship. The two-slope function that explained this dataset may not be robust to other datasets. None of the included studies used fewer than 29 sites; findings cannot be extrapolated below this threshold. The smallest study included only 142 subjects, which may make it prone to bias; however, the remainder had greater than 200 subjects, making this less of a risk.[28]

In summary, this analysis of opioid clinical trials suggests that the greater the number of sites, the smaller the observed effect size of treatment. This means that more subjects overall, and per site, need to be added to preserve power as the number of sites increases. Sample size calculations from smaller Phase 2 studies may seriously underestimate sample size requirements for larger Phase 3 studies. This may be one of several causes of failed Phase 3 clinical trials; conclusions should be interpreted in the context of the limitations of this study.

Disclosure

DSM is an employee of Collegium Pharmaceutical, Inc. BV is an employee of Rho. EAK is an employee at and owns stock with Teva Branded Pharmaceutical Products R&D, Inc. EAK was employed by Collegium with stock equity at the start of the study. NK is an employee of WCG Analgesic Solutions, a consulting and clinical research firm, and is affiliated with Tufts University School of Medicine. This work was not sponsored by, or conducted on behalf of, Collegium. The views expressed in this manuscript are those of the authors only and should not be attributed to Collegium. The authors report no other conflicts of interest in this work.

References

1. US Food and Drug and Administration (FDA), Center for Drug Evaluation and Research (CDER). *Guidance for Industry Analgesic Indications: Developing Drug Biological Products*. US Department of Health and Human Services; 2014.
2. US Food and Drug Administration (FDA). *Center for Drug Evaluation and Research (CDER). Guidance for Industry: Oversight of Clinical Investigations – A Risk-Based Approach to Monitoring*. Rockville, MD; 2013.

3. Dworkin R, Turk D, Peirce-Sandner S, et al. Considerations for improving assay sensitivity in chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2012;153:1148–1158. doi:10.1016/j.pain.2012.03.003

4. Marder S, Laughren T, Romano S. Why are innovative drugs failing in phase III? *Am J Psychiatry*. 2017;174(9):829–831. doi:10.1176/appi.ajp.2017.17040426

5. Katz N. Opioids: after thousands of years, still geting to know you. *Clin J Pain*. 2007;23(4):303–306. doi:10.1097/AJP.0b013e31803cb905

6. Katz N. Methodological issues in clinical trials of opioids for chronic pain. *Neurology*. 2005;65(12):131–139. doi:10.1212/WNL.65.12_suppl_4.S32

7. Meske D, Lawal O, Elder H, Langeberg V, Paillard F, Katz N. Efficacy of opioids versus placebo in chronic pain: a systematic review and meta-analysis of enriched enrollment randomized withdrawal trials. *J Pain Res*. 2018;11:1–12. doi:10.2147/JPR.S160255

8. Hale M, Tudor IC, Khanna S, Thipphawong J. Efficacy and tolerabilityof once-daily OROS hydromorphone and twice-daily extended-releaseoxycodone in patients with chronic, moderate to severe osteoarthritisspain: results of a 6-week, randomized, open-label, noninferiority analysis. *Clin Ther*. 2007;29(5):874–888

9. Katz N, Rauck R, Ahdieh H, et al. A 12-week, randomized, placebo controlledtrial assessing the safety and efficacy of oxymorphoneextended release for opioid-naive patients with chronic low back pain. *Curr Med Res Opin*. 2007;23(1):117–128

10. Hale M, Khan A, Kutch M, Li S. Once-daily. OROS hydromorphoneER compared with placebo in opioid-tolerant patients with chronic lowback pain. *Curr Med Res Opin*. 2010;26(6):1505–1518

11. Katz N, Hale M, Morris D, Stauffer J. Morphine sulfate and naltrexonehydrochloride extended release capsules in patients with chronicosteoarthritis pain. *Postgrad Med*. 2010;122(4):112–128

12. Friedmann N, Klutzaritz V, Webster L. Efficacy and safety of an extendedreleaseoxycodone (Remoxy) formulation in patients with moderate tosevere osteoarthritic pain. *J Opioid Manag*. 2011;7(3):193–202.

13. Rauck RL, Nalamachu S, Wild JE, et al. Single-entity hydrocodoneextended-release capsules in opioid-tolerant subjects with moderateto-severe chronic low back pain: a randomized double-blind, placebocontrolledstudy. *Pain Med*. 2014;15(6):975–985.

14. Wen W, Sitar S, Kynch S, He E, Ripa SR. A multicenter, randomized, double-blind, placebo-controlled trial to assess the efficacy andsafety of single-entity, once-daily hydrocodone tablets in patients withuncontrolled moderate to severe chronic low back pain. *Expert OpinPharmacother*. 2015;16(11):1593–1606

15. Katz N, Kopecky EA, O'Connor M, Brown RH, Fleming AB. A phase3, multicenter, randomized, double-blind, placebo-controlled, safety,tolerability, and efficacy study of Xtampza ER in patients with moderateto-severe chronic low back pain. *Pain*. 2015;156(12):2458–2467

16. Hale ME, Laudadio C, Yang R, Narayana A, Malamut R. Efficacy andtolerability of a hydrocodone extended-release tablet formulated with abusedeterrencetechnology for the treatment of moderate-to-severe chronic painin patients with osteoarthritis or low back pain. *J Pain Res*. 2015a;8:623–636.

17. Hale ME, Zimmerman TR, Eyal E, Malamut R. Efficacy and safety of ahydrocodone extended-release tablet formulation with abuse-deterrencetechnology in patients with moderate-to-severe chronic low back pain.*J Opioid Manag*. 2015b;11(6):507–518

18. Guthrie L, Oken E, Sterne J, et al. Ongoing monitoring of data clustering in multicenter studies. *BMC Med Res Methodol*. 2012;12:29. doi:10.1186/1471-2288-12-29

19. Kraemer H. Pitfalls of multisite randomzied clinical trials of efficacy and effectiveness. *Schizophr Bull*. 2000;26(3):533–541. doi:10.1093/oxfordjournals.schbul.a033474

20. Noda A, Kraemer HT, Schneider JL, Ashford B, Yesavage J. Strategies to reduce site differences in multisite studies: a case study of Alzheimer disease progression. *Am J Geriatr Psychiatry*. 2006;14(11):931–938. doi:10.1097/01.JGP.0000230660.39635.68

21. Spirito A, Abebe K, Iyengar S, et al. Sources of site differences in the efficacy of a multisite clinical trial: the treatment of SSRI-resistant depression in adolescents. *J Consult Clin Physchol*. 2009;77(3):439–450. doi:10.1037/a0014834

22. International council for harmonisation of technical requirements for pharmaceuticals for human use (ICH). ICH Harmonised guideline: integrated addendum to ICH E6(R1): Guideline for good clinical practiceE6(R2); 2016.

23. Bridge J, Iyengar S, Salary C, et al. Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. *JAMA*. 2007;297(15):1683–1696. doi:10.1001/jama.297.15.1683

24. Treister R, Lawal O, Schecter J, et al. Accurate pain reporting training diminishes the placebo response: results from a randomized, double-blind, crossover trial. *PLoS ONE*. 2018;13(5):e0197844. doi:10.1371/journal.pone.0197844

25. Mayorga A, Flores C, Trudeau JM, et al. A randomized study to evaluate the analgesic efficacy of a single dose of the TRPV1 antagonist mavatrep with osteoarthritis. *Scan J Pain*. 2017;17:134–143. doi:10.1016/j.sjpain.2017.07.021

26. Treister R, Eaton T, Trudeau J, Elder H, Katz N. Development and preliminary validation of the focused analgesia selection test to identify accurate pain reporters. *J Pain Res*. 2017;10:319–326. doi:10.2147/JPR

27. Edwards P, Shakur H, Barnetson L, Prieto D, Evans S, Roberts I. Central and statistical data monitoring in the clinical randomisation of an antifibrinolytic in significant haemorrhage (CRASH-2) trial. *Clin Trials*. 2014;11(3):336–343. doi:10.1177/1740774513514145

28. Lin L. Bias caused by sampling error in meta-analysis with small sample sizes. *PLoS ONE*. 2018;13:9.