American Pain Society
RESEARCH
EDUCATION
TREATMENT
ADVOCACY
PUBLISHED BY ELSEVIER

The Journal of Pain, Vol 14, No 8 (August), 2013: pp 767-777
Available online at www.jpain.org and www.sciencedirect.com

Critical Review

Integration of Pain Score and Morphine Consumption in Analgesic Clinical Studies

Feng Dai,* David G. Silverman,† Jacques E. Chelly,‡ Jia Li,§ Inna Belfer,‡ and Li Qin*

*Yale Center for Analytical Sciences, Yale University, New Haven, Connecticut.
†Department of Anesthesiology, Yale University, New Haven, Connecticut.
‡Department of Anesthesiology, University of Pittsburgh, Pittsburgh, Pennsylvania.
§Henry Ford Health System, Detroit, Michigan.

Abstract: **In pain clinical trials, the rescue analgesic medication such as patient-controlled analgesia morphine is often made available for patients for breakthrough pain. The patient-controlled analgesia morphine usage decreases the study agent's effect on pain relative to placebo and introduces greater variability in attainment of pain scores. For assessment of analgesic efficacy, the isolated statistical analysis of pain score or morphine consumption as a surrogate marker for pain not only loses statistical efficiency but also may incur increased false-positive findings because of multiple testing. The aim of this article is to review the research to date for choices of statistical tests for pain or morphine consumption outcome, with a focus on systematically evaluating a means for collective analgesic assessment of pain and morphine consumption using an integrated outcome. A case example is illustrated for data visualization, statistical comparison, and effect size estimation using the new endpoint. Some implications for clinical practice and further research are discussed.**

Perspective: ***This article provides statistical evidence to conclude that an integrated outcome of pain score and morphine consumption provides an efficient means for integrated analgesic assessment.***

Key words: *Pain clinical trials, analgesics, pain, PCA morphine, integrated analgesic assessment.*

The methods of postoperative pain assessment in many clinical studies changed when the use of patient-controlled analgesia (PCA) devices became standard practice in the 1980s. Clinicians and the institutional review board members became reluctant to deprive patients of these devices to conduct studies. Now, in most studies of analgesic agents and techniques, we must for ethical reasons ensure that rescue analgesia for the treatment of breakthrough pain is made available through PCA devices (in some studies, oral pain medicine may be offered).[28]

This work was supported by a career development grant from the U.S. Cancer Pain Relief Committee, awarded to Feng Dai.
None of the authors have any financial or other relationships that might lead to a conflict of interest.
Supplementary data accompanying this article are available online at www.jpain.org and www.sciencedirect.com.
Address reprint requests to Feng Dai, PhD, Yale Center for Analytical Sciences, Yale University, 300 George Street, Suite 555, New Haven, CT 06511. E-mail: feng.dai@yale.edu
1526-5900/$36.00

http://dx.doi.org/10.1016/j.jpain.2013.04.004

However, since the introduction of PCA, postoperative analgesic studies that use the pain intensity score outcome alone have lost efficacy because of the greater variability introduced by this method, because different patients often make different degrees of PCA rescue demands.[9,28] Therefore, in current clinical trials investigating the effectiveness of analgesic medications, instead of the traditional method's large decrease in pain, one is likely to see only a modest decrease in pain and a modest decrease in PCA dosing. This affects both study sensitivity and budget considerations, because a decrease of effect size means that the required sample size would increase accordingly.[36] In addition, assessing pain response and PCA use scores as 2 isolated variables not only may fail to identify the total benefit provided by the analgesic interventions being studied but also may increase the likelihood of reporting a false significant difference.[51]

Recent consensus meetings of analgesic researchers from academia, industry, and the United States Food

and Drug Administration[15,54] concluded that one of the highest priority tasks for analgesic study method research was to determine the optimal method for combining the information in patient pain scores with the amount of morphine they gave themselves through PCA pumps, and ultimately improving efficacy in clinical trials on novel analgesic interventions. In this regard, some innovations might even have the potential to turn some "failed new drug applications" into commercial successes.[36]

There was scant analgesic research focused directly on the topic; however, an early work by Silverman et al[51] appeared to be plausible to help recover the "lost power" of rescue dose paradigms. Instead of comparing pain or analgesic (morphine or equivalent) consumption separately, their method aimed at deriving an integrated analgesic assessment score for testing comparisons between analgesic treatments. There were a few early empirical findings that clinical implementations of the method showed promising results, with increased probability of detecting treatment effects. Unfortunately, no following work has been attempted to comprehensively evaluate the accuracy and effectiveness of the method in the context of pain clinical trials where rescue analgesia is available to all patients.

Therefore, our aim is to review the research to date for choices of statistical tests for pain or morphine consumption outcome, with a focus on systematically evaluating a means (ie, Silverman integrating approach [SIA]) for collective analgesic assessment of pain and morphine consumption using an integrated outcome. The outline of the paper is as follows: First, we review common choices of statistical tests used in analgesic or pain clinical trials. Second, we review statistical methods for global evaluation of analgesic efficacy highlighting the fine statistical property of SIA by numerical Monte Carlo simulations. Third, we apply the SIA method to a real data example, illustrating why the SIA provides a reliable framework for data visualization, statistical comparison, and effect size estimation for jointly analyzing pain and morphine consumption outcomes. Last, we conclude the paper by discussing the implications of our findings for clinical practice and further investigation.

Survey of Methods for Analyzing Pain and Morphine

We will discuss 2 types of statistical analysis methods that have been utilized for pain and analgesic study: separate analysis and global evaluation of analgesic efficacy.

Separate Analysis of Pain and Morphine

Since the use of PCA devices became standard practice, the effect of an analgesic is now at least divided between pain scores and PCA morphine usages. Different approaches have been used to specify a clinical decision rule for clinical trials that have 2 or multiple endpoints.[16] The most common strategy is to choose a single outcome (eg, pain intensity score), with all others specified as secondary outcomes (eg, morphine usage as a pain surrogate marker).

Depending on the distribution of outcome, a parametric Student t-test or a nonparametric Wilcoxon-Mann-Whitney (WMW) rank sum test has often been used to compare 2 unrelated groups, with parametric analysis of variance (ANOVA) or nonparametric Kruskal-Wallis test being used to compare 3 or more groups. In general, these nonparametric tests are more powerful than their parametric counterparts when the underlying distributions are departed from normality with heavy-tails (ie, high skewness). When data are normally distributed, nonparametric tests are less powerful; however, the power loss is not substantial especially when sample size is not small.[26]

In randomized analgesic or pain clinical trials, empirically a variety of statistical test methods such as ANOVA, Kruskal-Wallis test, t-test, and WMW rank sum test were used to analyze pain or rescue analgesia outcome.[2,5,12,20,21,25,30,31,45,46,48,53,56] Two extensive simulation works had been performed by Dexter[10,11] to compare in particular the accuracy and effectiveness of several statistical tests mentioned earlier to detect differences in pain or analgesic usage among groups. It was found that the choice of an appropriate test always depends on the outcome of interest to be compared among groups. On the one hand, as doses of analgesic measurements often are highly skewed, the WMW rank sum test and Kruskal-Wallis test are found to be the best tests to detect the differences among groups.[10,40] On the other hand, it was found that as pain measurements like the visual analog scale[14,8] generally follow closer to a normal distribution, the parametric t-test and ANOVA have higher statistical powers and therefore are more appropriate methods to detect differences among groups.[11]

Global Evaluation of Analgesic Efficacy

Certain procedures are available for testing the global null hypothesis of no treatment effects where the vector of outcomes has a multivariate normal distribution. For example, the parametric Hotelling's T^2 test, a multivariate analog of the univariate t-test, could be used to compare the jointed difference of 2 or more endpoints in 2 treatment groups. However, its strong distributional assumption of multinormality is often doubtful for real data (eg, the analgesic consumption is usually skewed or zero-inflated), particularly if data sample sizes are small. In addition, Hotelling's T^2 test is sensitive to treatment effects that are opposite in sign for different endpoints (ie, the treatment has beneficial effect on one outcome but detrimental effect on the other), which limits the use of this procedure.

Silverman et al[51] derived an integrated analgesic assessment score (SIA score) by combining the transformed ranking values of pain and morphine use