Moving towards multiple site outcomes in spinal cord injury pain clinical trials: An issue of clustered observations in trial design and analysis

Elizabeth Richardson¹, David T. Redden²

¹Department of Physical Medicine and Rehabilitation, ²Department of Biostatistics, The University of Alabama at Birmingham, Birmingham, AL, USA

Introduction: Pain remains a problem for many with spinal cord injury (SCI), and there is a need for sound, randomized clinical trials examining the efficacy of existing and novel therapeutics. SCI-related pain is complex, as more than one type of pain is often experienced. The purpose of this report is to (i) demonstrate how to design and power calculation of a clinical trial of SCI pain using multiple pain sites per individual; (ii) discuss consequences of failing to adjust for this; and (iii) provide intraclass correlation (ICC) estimates for common pain outcome measures that may be used to power future clinical trials in SCI pain.

Method: Using an existing dataset from a past SCI pain clinical trial, the ICC was calculated for common pain outcome measures to illustrate appropriate corrections for powering, analyzing and interpreting results from multiple pain sites per individual. The problem associated with not accounting for multiple pain sites per individual and the effect on the Type I error rate is also shown.

Results and Discussion: Not accounting for the ICC can lead to (1) incorrect power estimates in the design of a trial, and (2) an inflated Type I error rate with a higher likelihood of misinterpretation of outcomes.

Conclusions: Powering for future SCI pain trials and statistical analysis of trial outcomes may be substantially compromised if methods do not account for the intra-individual associations between pain sites, ultimately affecting study interpretations and evidence-based practice. We present ICC estimates based on SCI pain data for purposes of estimating power for future trials.

Keywords: Spinal cord injury, Pain, Clinical trials, RCT, Intraclass correlation coefficient, ICC

Introduction

Persistent pain is a common secondary condition following spinal cord injury (SCI). Approximately 70% of persons with SCI report pain, with a third of those describing the pain as severe.¹ Functional impairments arising from pain are pervasive, disrupting mood and psychosocial functioning,^{2,3} occupational activities,⁴ and basic needs such as sleep.^{5,6} Thus, it is not surprising that pain, among other complications of SCI, is consistently associated with lower quality of life post-injury.⁷⁻⁹

Currently available treatment for SCI-related pain includes pharmacological substances (e.g. gabapentin,

pregabalin), surgical procedures such as dorsal root entry zone ablation, physical therapies, and psychological interventions, although Levels I and II empirical evidence in support of the current clinical armamentarium remains limited.¹⁰⁻¹³ Moreover, SCI-related pain, and in particular neuropathic pain (NP), has proven modestly to minimally responsive to a variety of pharmacological and non-pharmacological treatments,¹⁴⁻¹⁷ leaving many individuals with SCI experiencing refractory pain despite currently available treatment.¹⁸⁻²⁰ Clearly, there is a great need in the field of SCI-related pain to conduct sound, randomized clinical trials of not only currently available treatments but also novel therapeutics developed at the bench to minimize futility and increase confidence that the clinical interpretation of trial outcomes are accurate.

Correspondence to: Elizabeth J. Richardson, Department of Physical Medicine and Rehabilitation, The University of Alabama at Birmingham, 1717 6th Avenue South, Birmingham, AL 35249-7330, USA.
Email: ejrichar@uab.edu

A complicating issue in designing clinical trials in this area is the complex nature of pain following SCI. Pain following SCI is not a unitary phenomenon; there are in fact several subtypes of pain, each with different purported mechanisms and thus assumed to be differentially responsive to various treatments. Most SCI-related pain taxonomies broadly classify SCI-related pain into two types: musculoskeletal (MS)/nociceptive and neuropathic, and indicate where, relative to the location of the injury, the particular subtype of pain is experienced.^{7,21} In addition, one or more subtypes of SCI-related pain can be present in multiple areas of the body.

To this end, the International Spinal Cord Injury Basic Pain Data Set (ISCIBPDS)²² proposed standardization of collecting pain outcome data among persons with SCI in order to promote better interpretation and translation of clinical outcomes. Based on the recommendations of the Initiative on Methods, Measurement, and Pain assessment in clinical Trials (IMMPACT),²³ the ISCIBPDS includes a core set of outcomes that reflect the multiple dimensions that comprise the construct of pain while also accounting for the complexity and potentially multiple locations of SCI-related pain. To accomplish this, the ISCIBPDS working group recommended that up to the three worst pain sites should be assessed per individual via pain severity ratings and other core pain measures.²²

Since the ISCIBPDS recommendations were outlined in 2008²² and at the time of writing this article, 15 prospective clinical trials in SCI-related pain have been published and indexed in Medline®. While one intervention trial examined heterogeneous chronic SCI pain,²⁴ the majority of trials focused on neuropathic SCI pain specifically but did not differentiate between subtypes of NP.^{25–31} In other studies, if individuals experienced multiple NP sites, severity ratings from either the worst site or the average across all sites were calculated to produce a single unit of analysis.^{32–34} A pilot trial of dronabinol for SCI NP controlled for NP subtype by analyzing only below-level NP,³⁵ while two other trials either differentiated subtypes of SCI NP³⁶ or stratified pain sites based on neuropathic descriptors (mechanical allodynia, dysaesthesia, paroxysmal, or continuous).³⁷ However, for these last two studies, we could not identify analytic approaches that would be appropriate for non-independence or clustering of pain outcomes (e.g. individuals contributing data to both at- and below-level pain, multiple pain subtypes, or endorse multiple symptom categories within a subtype). This is not to say that the inferences made from these studies are not valid; rather, it is to suggest that the methodology was unclear as to how multiple pain types/sites per

individual were handled. Aside from a study we have recently published,³⁸ only one other trial utilized the ISCIBPD methods and collected outcome data on multiple pain types and sites³⁹; unfortunately, the small sample size of this latter study precluded full analysis and these data were examined only descriptively.

While averaging across pain sites to produce one outcome or simply focusing on a single subtype and location of pain is appropriate, accounting for multiple pain sites provides the obvious benefit of investigating differential effects of interventions on subtypes of SCI pain. To this date and to our knowledge, no group has published a discussion with regard to the issues of both design and analysis of a clinical trial utilizing multiple pain sites per individual with SCI in accordance with the ISCIBPD.²²

Observations from more than one pain site per individual with SCI will correlate and thorough discussions of the consequences of such a correlation, called the intra-class correlation coefficient (ICC), exist within the statistical literature.^{40–42} Clustered observations are quite common in epidemiological research (longitudinal research), medical research (patients nested within hospitals or practices), dental research (teeth nested within a person), and educational research (students nested within schools). These fields use similar statistical methods to account for the correlation among nested observations in analysis; however, the degree of correlation can vary greatly depending on what constitutes a “cluster”, leading to significant problems when designing a clinical trial. Hierarchical linear models, multilevel models, random effects models, mixed effects models, and mixed linear models are terms used across disciplines to describe the same statistical method of controlling for the effects of clustering. For the purpose of this report, we will use the term *mixed linear modeling*.

As the need grows for high-level clinical trial research in SCI pain, the overarching purpose of this present report is to offer tools for researchers in the field of SCI pain to maximize validity of outcomes that are ultimately translated to clinical practice. Within this report, we demonstrate how to design and power a clinical trial of SCI pain using multiple pain sites per individual by calculating and accounting for the ICC that we have calculated from common pain outcomes used in SCI research. We will also discuss the impact of the ICC in estimating the variance of clustered SCI pain outcomes and consequences of failing to adjust for this factor in post-trial analyses. Lastly, we provide ICC estimates for common pain outcome measures that may be used by others as point estimates in power calculations for future studies. We present these concepts utilizing a

previously collected dataset that, in accordance with ISCIBPD methodology, recorded outcomes for multiple pain sites per person with SCI.

Methods

Appropriate methods of design and powering SCI pain trials

Calculation of the ICC coefficient

The simplest definition of an ICC is the proportion of the outcome variance that is attributable to clustering. Within pain studies, clusters are simply the individuals (represented below by the subscript i , where $i = 1$ up to n clusters or individuals) and pain sites are observations nested within an individual (represented by j , where $j = 1$ up to m_i observations or pain sites). Within this paper, the individual represents the cluster and observations are the pain sites nested with the individuals. For the remainder of this paper, when we reference the cluster level we will simply use the term “individuals”. Note the number of pain sites m_i vary by individual indicating a different number of pain sites can be observed for each individual. A continuous outcome, such as pain score, can be represented as:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (1)$$

This simple equation illustrates that the outcome, Y_{ij} , can be conceptualized as a function of a grand mean (population average score), a random effect that varies by individual (α_i), and a random error component (ϵ_{ij}). In studies of clustered observations, the “grand mean” is simply the average of the all observations collected. The term α_i represents the unique effect of the i th individual and is assumed to be normally distributed with mean = 0, variance = σ_α^2 . This term, σ_α^2 , accounts for the observed variation among individuals. The term, ϵ_{ij} , is also a random variable that accounts for variation within individuals and is assumed to be normally distributed with mean = 0 and variance = σ_ϵ^2 . The two random effects are assumed to be independent; thus, the total variance of the outcomes (Y_{ij}) is $\sigma^2 = \sigma_\alpha^2 + \sigma_\epsilon^2$. To estimate the proportion of the outcome

variance due to variation among individuals, one defines ICC as follows:

$$\frac{\sigma_\alpha^2}{\sigma^2} = \frac{\sigma_\alpha^2}{(\sigma_\alpha^2 + \sigma_\epsilon^2)} \quad (2)$$

where by definition in equation 2, the ICC must be between 0 and 1.

All these components are equivalent to terms included in random-effect analysis of variance (ANOVA) and therefore can be estimated by mean squares between (MSB) and mean squares error (MSE) from an ANOVA table. Therefore, each component can be estimated by running an ANOVA, where the between groups source of variance is the unit by which the clusters are measured. Murray⁴⁰ illustrates that the ICC can be estimated as:

$$\frac{\text{MSB} - \text{MSE}}{\text{MSB} + (m' - 1)\text{MSE}} \quad (3)$$

Table 1 demonstrates the components of the variance. Here, n is the number of individuals and, because of clustering, is analogous to the number of “groups”. N is the total number of pain sites, and therefore, $N = \sum_{i=1}^n m_i$.

Use of the adjusted mean cluster size will more accurately estimate the variability attributable to the individual when the cluster sizes are not equal.⁴³ In real data involving SCI pain sites, it will be common for individuals (clusters) to have varying numbers of pain sites (observations). Because of this issue, use of the adjusted mean cluster size, rather than the simple mean cluster size, is recommended. The adjusted mean cluster size and can be calculated via the following:

$$m' = \frac{1}{(n - 1)} \left[N - \sum_{i=1}^n \frac{m_i^2}{N} \right] \quad (4)$$

Following the computation of the adjusted mean cluster size, it is possible to simplify estimation of the ICC as a function of the F statistic from a computed ANOVA

Table 1 Sources of variance as estimated by random-effect ANOVA

Source	Sums of squares	Degrees of freedom	Mean squares	F ratio
Between	$\text{SSB} = \sum_{i=1}^n \frac{(\sum_{j=1}^{m_i} Y_{ij})^2}{m_i} - \frac{(\sum_{i=1}^n \sum_{j=1}^{m_i} Y_{ij})^2}{N}$	$n - 1$	$\text{MSB} = \frac{\text{SSB}}{(n - 1)}$	$F = \frac{\text{MSB}}{\text{MSE}}$
Error	$\text{SSE} = \sum_{i=1}^n \left(\sum_{j=1}^{m_i} (Y_{ij} - \frac{\sum_{j=1}^{m_i} Y_{ij}}{m_i})^2 \right)$	$N - n$	$\text{MSE} = \frac{\text{SSE}}{(N - n)}$	

SSB, sum of squares between; SSE, sum of squares error; MSB, mean squares between; MSE, mean squares error.

table, like that which is shown in Table 1. In this manner, the ICC can be computed via the following equation, and as illustrated further in the Results section.

$$\frac{(F - 1)}{(F - 1 + m')} \quad (5)$$

Calculation of the variance inflation factor

The consequence of a positive ICC among observations nested within an individual is an inflation of the variance due to non-independent observations that must be accounted for within the power calculation. Specifically, this increase in variance is captured by the following formula for the variance of clustered observations:

$$\sigma^2 + \sigma^2(m' - 1)ICC = \sigma^2[1 + (m' - 1)ICC] \quad (6)$$

The multiplicative factor in the second equation of equation 6 above is often referred to in literature⁴¹ as a variance inflation factor (VIF) and shows that the variance of clustered observations can be conceptualized as two components: the variance that would occur if one observation was randomly sampled from the m' observations available from the cluster, (σ^2), and the inflation due to utilizing the remaining $m' - 1$ observations. For example, if a researcher was anticipating using an outcome measure that was psychometrically designed to have a variance of 10 points but also anticipated recording an adjusted average of 3 observations per subject enrolled with an ICC of 0.5, the researcher should not conduct the power analysis assuming a variance of 10 points. Given the clustered nature of the data, the researcher should anticipate a variance of 20 points, or $10 \times (1 + (3 - 1) \times 0.5)$, twice the variance that would be anticipated for non-clustered data. From this equation, it can easily be seen that the consequence of increasing ICC values is an increasing inflation in variance estimation.

Adjustment of power calculations to account for ICC

As illustrated previously,^{40,42} estimating the sample size required for a proposed study is easily achieved by modifying traditional sample size calculations to adjust for the VIF. Under the traditional approach, the researcher is expecting one observation per subject. Independence is correctly assumed to determine the number of individuals needed per group, n_{ind} . To calculate a needed sample size when a study design requires multiple observations from each individual, n_{clus} , use of the VIF and adjusted mean cluster size, m' , will correctly adjust for

the expected inflation in the variance:

$$n_{clus} = \frac{n_{ind} \times VIF}{m'} \quad (7)$$

Therefore, one would need a total of n_{clus} persons per group, with each individual providing an average of m' observations to achieve the desired statistical power. By this calculation, a researcher should expect a total number of outcomes, or pain sites per group to be $n_{clus} \times m'$.

Appropriate methods of SCI pain trial data analysis

Just as researchers must account for the ICC in the design of the clinical trial or experiments that uses multiple pain sites per subject, researchers must account for the ICC in the analysis. The approaches of two-sample t -tests, linear regression, logistic regression, and tests of proportions must all account for the extra variation attributable to the ICC. Numerous fields have addressed this issue with methods such as fixed effects regression, robust cluster variance estimation, mixed linear models, generalized linear mixed models, and generalized estimating equations (GEEs). Therefore, this paper focuses more on mixed linear models, generalized linear mixed models, and GEEs, which are closely related to robust cluster variance estimation.

For continuous outcomes, the following mixed linear model is appropriate for multiple observations per subject.

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \alpha_i + \varepsilon_{ij} \quad (8)$$

The above equation indicates two important facts. First, characteristics that vary by pain site (such as pain type, symptoms, or location) can be included in the equation by covariate X_{1ij} (the explanatory variable for the i th person at his or her j th pain site). Second, person-level characteristics, such as gender, race, injury etiology, and smoking status, which do not vary by pain site, can also be included in the model. This type of variable is denoted by X_{2i} , indicating the explanatory variable does not vary by pain site for the i th individual. Good reviews of mixed linear models are presented in Raudenbush and Bryk⁴⁴ and Brown and Prescott.⁴⁵

With regard to dichotomous outcome variables, a similar model can be built using generalized linear mixed models. Defining the probability that $Y_{ij} = 1$ is π_{ij} , then the generalized linear mixed model is written as:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \alpha_i \quad (9)$$

In the equation above, α_i again represents a random effect that is attributable to the i th individual. These terms account for the variance among the individuals, with the remaining terms of the equation simply representing the standard logistic regression notation. Another approach similar to mixed linear models and generalized linear mixed models is GEEs.⁴⁶ Although GEEs do not directly model the random effects, the approach adjusts variance estimates for clustering. It should be noted that mixed linear models, generalized mixed linear models, and GEEs maintain the correct Type I error rate only with a sufficient number of individuals (clusters).

Whereas both mixed linear models and generalized linear mixed models assume a specific distribution for the random effects, another popular approach for accounting for the ICC, called GEEs, does not specify the distribution of the random effects and is therefore suitable when the outcome of interest is dichotomous. Instead, GEE models provide estimates of the regression coefficients and their variances adjusted for the ICC by making assumptions about the pattern of correlations among observations nested within a cluster. This pattern of correlations among observations within a cluster is called the working correlation structure. It has been shown that GEE models, given a large number of clusters, provide consistent estimates of variance even if the working correlation structure is incorrectly specified. When compared with robust variance estimation techniques, GEE demonstrated comparable results with regard to Type I error control and statistical power.⁴⁷

Data used in example power calculation and analyses

Data from a previously conducted randomized, double-blind clinical trial examining effects of nicotine on SCI-related pain³⁸ were used to estimate ICC values for common pain outcomes and to outline the suggested analytic methods presented in this report. For this dataset, the total number of individuals with SCI (n) was 42. The sample was 52.4% African American (47.6% Caucasian) and 38.1% women (61.9% men). Approximately 35.7% had tetraplegia (64.3% with paraplegia) and 78.6% had incomplete injuries (21.4% with complete injuries).

Consistent with recommendations from the ISCI-PBDS,²² outcomes from a maximum of three worst pain types per subject were assessed and available in the dataset. Two interviewers, trained to use the Bryce–Ragnarsson pain classification scheme¹⁹ simultaneously but independently classified individuals'

pain sites. Pain sites were classified as MS, pure NP, or some combination of MS and NP symptoms (complex neuropathic). This resulted in 42 (40.78%) MS sites, 39 (37.86%) NP sites, and 20 (19.42%) complex NP sites in the study. Two people were found to have at least one pain site that met criteria for visceral pain (1.94%). In total, the number of pain sites (N) was 103; however, given the small number of visceral pain sites, visceral pain was neither included in the primary analyses of the previous trial³⁸ nor in the example analyses and illustrations presented here. Although the adjusted mean cluster size, m' , varies slightly across the outcomes, we have chosen to use $m' = 2.45$ in this report.

It has been suggested that pain, as a core outcome domain within a clinical trial, consists of multiple dimensions that should be assessed.⁴⁸ These dimensions include intensity, specific verbal descriptors, and quality of pain symptoms.⁴⁸ The trial from which these data were derived was consistent with IMMPACT recommendations,²³ and used a 0–10 numeric rating scale (NRS) to measure pain intensity and the short form McGill Pain Questionnaire (SF-MPQ)⁴⁹ to measure specific sensory pain qualities and the affective component of pain. The SF-MPQ consists of 15 verbal descriptors of pain, 11 of which represent sensory dimensions of pain with the remaining four representing affective qualities of pain. For use among persons with SCI, two sensory descriptors, “tingling” and “numbness”, from the long-form MPQ, are often included among the sensory descriptors,^{50,51} as was done here. The SF-MPQ sensory and affective scores can be summed to obtain a total score. Pain outcome data from the placebo condition were used to conduct example calculations in this report.

In addition to use of this dataset for purposes of illustrating potential problems associated with design and analysis of clustered SCI pain data, data simulations were also conducted to examine the effect of not accounting for the ICC on Type I error rate in statistical interpretation of results. To accomplish this, we simulated the condition of a true null hypothesis (no effect between conditions) with data parameters similar to that which was found in our SCI pain trial.³⁸ We allowed for intraclass coefficient values to range from 0 to 0.5, incremented by 0.10. One thousand simulations were conducted per ICC value, where each simulation generated two groups each with 100 individuals and 3 pain sites per individual. All simulations were set up such that the expected means of the two groups were equal; hence, if a P value of 0.05 or less was observed, then a Type I error had occurred. To compare the

Type I error rates for an incorrect approach (linear regression) to the Type I error rates for an appropriate procedure (mixed linear models), hypothesis tests were conducted for each simulation at an ICC value. For each level of ICC, the Type I error rate as a function of ICC was recorded and graphed. All simulations were conducted using SAS 9.3.

Results and discussion

This section will report our outcomes when the tools described above are applied to our SCI trial dataset consisting of multiple pain sites per individual. In this process, the pitfalls of failing to account for multiple pain sites per individuals with SCI in the design and analysis will also be highlighted. We begin with the calculation of the ICC for the present NRS score. Specifically, an ANOVA model was constructed, where each of the 42 individuals represented was coded as a unique “group” and pain sites ($N = 103$) were treated as observations within an individual. This calculation led to an estimated $MSB = 14.42$, $MSE = 2.46$, $m' = 2.45$, and $F = 5.85$. Using equation 3 or 5, the ICC for present NRS is 0.664 indicating a large degree of clustering for pain rating. Table 2 shows the ICCs and VIFs calculated for several pain outcomes used in SCI pain trials that are also included in the recommendations made by IMMPACT.²³

Example power calculations for a trial

Continuous outcomes

To illustrate how the below ICC estimates and VIF values can be used to calculate power for a two-sample t -test approach, we will assume a study is being designed to investigate the outcome as measured by the SF-MPQ sensory subscale. Based upon literature,⁴⁹ one could expect a standard deviation of 8.7, which equates to variance of 75.69. Assuming that a

difference of three points between the two group means is assumed to be clinically relevant, standard power calculation software indicate that 133 individuals are needed per group, for a total sample size of 266 to achieve 80% power to detect that difference and assuming a Type I error of 0.05. However, this calculation assumes only one pain site per subject is observed. According to equation 9, if an average of 2.45 pain sites per individual is expected, the sample size should be inflated by 1.978. The adjustment is simply $133 \times 1.978 = 263.07$ to arrive at an estimated requirement of 264 observations (pain sites) per arm, or 528 observations (pain sites) in total. The number of *individuals* or “clusters” to recruit per group, is the total number of observations, 264, divided by the adjusted cluster size, 2.45. This results in a need for 107.75 or 108 individuals with SCI per group to achieve 80% power to detect a three-point difference between group means and assuming a Type I error rate of 0.05, 2.45 pain sites per subject, an ICC of 0.674, and an assumed variance of 75.69.

With the ICC information provided in Table 2, assume a study is being designed to investigate SF-MPQ affective outcome. From our data, we estimated the ICC to be 0.893. If one expects a different cluster number, such as two pain sites per individual, then one should anticipate a VIF of 1.893 (using the multiplicative factor, $(1 + (m' - 1)ICC)$, shown in equation 6). Finally, if an effect size of 0.5 (difference of the group means divided by a common standard deviation) is clinically relevant, then 64 individuals per group are required to achieve 80% power using a two-tailed Type I error of 0.05 and assuming one pain site per person. To adjust for two pains sites per person, a researcher would need to obtain $64 \times 1.893 = 121.15 = 122$ observations, which equates to 61 individuals per arm.

Dichotomous outcome

The same approach can be used for powering a study designed with dichotomous outcomes. Assume that a researcher seeks to implement a treatment that would reduce the probability of allodynia for a given pain site from 0.30 to 0.15. Table 2 shows that the ICC for allodynia is estimated to be 0.137. Not accounting for clustering, 121 individuals per group are needed to achieve 80% power to detect the difference in probabilities of 0.30 and 0.15, using a two-tailed type error rate of 0.05. In order to account for multiple pain sites per subject and assuming three pain sites per person, the researcher should expect a VIF of $(1 + (m - 1)ICC) = (1 + (2 \times 0.137)) = 1.274$. To adjust the

Table 2 ICC and VIF estimates for continuous and categorical pain outcomes

SF-MPQ scales	F value	ICC value	VIF
Sensory	6.07	0.674	1.978
Affective	21.53	0.893	2.295
Total	9.00	0.766	2.110
PPI	4.43	0.583	1.846
Numeric rating for severity			
NRS, present	5.85	0.664	1.963
NRS, average (24 hours)	6.15	0.678	1.983
NRS, average (week)	3.71	0.525	1.762
Bivariate outcomes			
Pain is allodynic (yes/no)	1.39	0.137	1.199
Pain is hyperalgesic (yes/no)	1.17	0.065	1.094

ICC, intraclass correlation; VIF, variance inflation factor; NRS, numeric rating scale; PPI, =present pain intensity.

sample size to account for clustering of three pains sites per person, $121 \times 1.274 = 154.15$ or 155 observations per arm are needed, which equates to 51.67 individuals per arm or 52 people per group.

Example data analysis of trial outcomes

Continuous outcomes

To illustrate building a mixed linear model accounting for multiple pain sites per subject, we tested an assumed hypothesis that etiology of SCI (gunshot vs. non-gunshot) affects pain outcomes, with those injured via gunshot having higher pain severity. For this example, we use the following model:

$$Y_{ij} = \beta_0 + \beta_1 X_{1i} + \alpha_i + \varepsilon_{ij}$$

where Y_{ij} is the SF-MPQ total score for the j th pain site nested within the i th subject, X_{1i} is an indicator of the person-level variable etiology (gunshot vs. non-gunshot) for the i th individual, α_i is the random effect for the i th individual, and ε_{ij} is the residual. In this model, β_1 measures and provides a test of whether SF-MPQ scores are affected by etiology. If one incorrectly analyzed the data, i.e. treating all 103 pain sites as independent observations, results would indicate that pain ratings for individuals injured by gunshot were, on average, 4.34 points higher compared with individuals injured by non-violent means. This incorrect approach would indicate significance with a P value of 0.02. However, correctly modeling the data leads to a non-significant difference ($P = 0.16$), with pain ratings for individuals injured by gunshot being 3.63 points higher compared with individuals injured by non-violent means.

Dichotomous outcomes

To illustrate the correct analysis of multiple pain sites per subject when the outcome is dichotomous, we tested the hypothesis that an individual's age is associated with pain severity. The numerical rating scale (0–10) for each pain site was dichotomized into severe pain (rating ≥ 7) or not severe pain ($0 < \text{rating} < 7$). In a logistic regression model which does not account for the multiple pain sites per individual, the regression estimates reveal that for every 1-year increase in the subject's age, the odds of severe pain multiplies by 1.06 with a 95% confidence interval ranging from 1.01 to 1.12. If one chooses to examine the P value from the analysis ($P = 0.015$), one would conclude a significant association between the log odds of severe pain and age, yet there is a higher likelihood that the null is in fact true despite the significant test. When a generalized linear mixed model is conducted, introducing a random

component for individual and therefore accounting for multiple sites within an individual, then the estimated odds ratio of 1.07 is deemed non-significant ($P = 0.095$) using a Type I error rate of 0.05. A 95% confidence interval for the odds ratio ranges from 0.98 to 1.16.

Is accounting for the ICC necessary?

Data simulations

The above example analyses illustrate a critical point about analyzing clustered or nested observations: the resulting P values may lead the researcher to different conclusions depending upon whether the clustering of pain sites is accounted for. If incorrect methods are applied, the probability of a Type I error increases as the ICC increases in magnitude. In order to illustrate the danger of not accounting for ICC, a simple simulation was conducted. As can be seen in the Fig. 1 below, when the ICC = 0 and pain sites within individuals are not correlated, both methods provide an appropriate Type I error rate. However, as the ICC increases, the incorrect method's Type I error rate inflates quickly, while the mixed linear model approach maintains the correct level.

When multiple observations within an individual are correlated, then this violates one of the major assumptions of hypothesis testing and standard statistical procedures (e.g. ANOVA, two-sample t -tests, logistic regression, multivariable linear regression, general linear models) are not appropriate to test hypotheses for drug or intervention effects. It is well documented^{40–46} that ignoring the ICC and treating all observations as independent leads to misspecification of the statistical distribution and inflates Type I error rate for hypothesis testing. This problem is illustrated with the data simulations presented in this report, which

Figure 1 Type I error rates as a function of ICC value.

modeled parameters likely to be encountered in SCI pain trials examining multiple pain sites per individual. In the trial dataset, we found ICCs from the previously conducted SCI pain trial to be quite large and commonly 0.5 and above. Results from the simulations we conducted suggests that when one does not account for the correlation of pain sites within an individual with SCI, a researcher may falsely reject the null hypothesis >20% of the time, when in fact the null hypothesis is true.

Other methodological considerations

Given the problems that may arise when failing to account for multiple pain sites per individual and the increased complexity when modeled, why not simply average across pain sites? If a researcher is interested in person-level or subject-specific explanatory variables, then averaging across pain sites to produce a single outcome for each individual is a viable and appropriate approach. Averaging across pain sites is only valid if, for instance, the question in study pertains to only one type of pain (e.g. neuropathic). Even under this condition, however, averaging across pain sites might be inappropriate because site-specific explanatory variables may vary by location. For example, when a specific type of pain is under question, such as SCI NP, a lack of distinction between at- and below-level SCI NP may confound results given that these two types of NP may have different underlying pathophysiological mechanisms and therefore show a differential response to treatment.⁵² In our recent trial on the effects of nicotine on SCI pain, collapsing across all pain sites that exhibited neuropathic symptomatology would have likely attenuated the significant effect occurred when types of NP were differentiated.³⁸

Another statistical approach that could be used, under limited circumstances, is repeated-measure ANOVA assuming an exchangeable or compound symmetric correlation matrix among pain sites. The assumption in this case is equivalent to the standard repeated-measure ANOVA assumption of sphericity. Whereas mixed linear models often assume this structure, repeated-measure ANOVA requires an *equal* number of pain sites per subject, which is very unlikely to be encountered in a natural sample of persons with SCI pain. Due to this restriction, most computer software programs will remove all patients with any number of pain sites less than the maximum observed. This fact vastly limits the utility of repeated-measure ANOVA in spinal cord pain trials.

Another approach commonly recommended is fixed effects regression. Within fixed effects regression,

variation among individuals is directly modeled by inclusion of indicator variables for each individual. A limitation of fixed effects regression is the prohibition of including other person-level variables⁵³ such as level of injury, injury severity, duration of injury, smoking status, sex, or race.

Summary and conclusions

Imperative to the advancement of the field of SCI pain is not only high-level, randomized controlled trials of pharmacological or non-pharmacological treatments, but also the confidence that results from such trials are valid when translated to clinical practice. SCI-related pain is a quite complex phenomenon, often representing a heterogeneous group of subtypes even within a single broader category (e.g. at-level neuropathic vs. below-level NP). Whereas studying multiple pain sites per subject increases the complexity of design and analysis of pain clinical trials, doing so will produce a more refined, rigorous, and thorough investigation to improve our understanding of underlying mechanisms and potentially different responses to treatment when pain site-specific factors are taken into account.

Proper design of clinical trials and subsequent interpretation of trial outcomes utilizing the ISCIBPDS methodology will be compromised if the design and analysis of pain data do not account for the intra-individual associations among pain sites within an individual. Ignoring this intra-individual association, estimated by the ICC, leads to uncertainty about whether variability in outcomes is attributable to the research effect in question or an individual factor that may lead to all pain sites for a given individual being rated higher/lower than comparable sites for another individual.

When designing a trial in accordance with the ISCIBPDS,²² in which multiple pain sites per individual are assessed, standard power calculations will not provide valid estimates of needed sample sizes to detect treatment effects. In order to account for correlation among pain sites within an individual, ICC coefficients must be either assumed or estimated in order to ensure adequate sample sizes to provide methodological validity and accuracy of results. Using the ICCs estimated within this report as guide, we recommend conducting power calculations over a range of assumed ICC values in order to estimate the sensitivity of the proposed sample size to variations in the ICC assumption. The ICCs estimated within this report should be considered point estimates of ICC and values around these estimates should be examined in designing future studies.

Acknowledgment

Data used to illustrate design and analysis issues in this report were originally collected for a clinical trial that was funded by the National Institute on Disability and Rehabilitation (NIDRR) grant no. H133N060021. Support for development of the manuscript was also provided by National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) grant no. P60AR048095-06A1.

References

- Bonica JJ. Introduction: semantic, epidemiologic, and educational issues. In: Casey KL (ed.) Pain and central nervous system disease: the central pain syndromes. New York: Raven Press; 1991. p. 13–29.
- Richards JS, Meredith RL, Nepomuceno C, Fine PR, Bennett G. Psychosocial aspects of chronic pain in spinal cord injury. *Pain* 1980;8(3):355–66.
- Elliot TR, Harkins SW. Psychosocial concomitants of persistent pain among persons with spinal cord injuries. *Neuro Rehabil* 1991;1(4):7–16.
- Rose M, Robinson J, Ellis P, Cole J. Pain following spinal cord injury: results from a postal survey. *Pain* 1988;34(1):101–2.
- Budh CN, Hultling C, Lundeberg T. Quality of sleep in individuals with spinal cord injury: a comparison between patients with and without pain. *Spinal Cord* 2005;43(2):85–95.
- Widerström-Noga EG, Felipe-Cuervo E, Yezierski RP. Chronic pain after spinal injury: interference with sleep and daily activities. *Arch Phys Med Rehabil* 2001;82(11):1571–7.
- Richards JS. Spinal cord injury pain: impact, classification, treatment trends, and implications from translational research. *Rehabil Psychol* 2005;50(2):99–102.
- Putzke JD, Richards JS, Hicken BL, DeVivo MJ. Interference due to pain following spinal cord injury: important predictors and impact on quality of life. *Pain* 2002;100(3):231–42.
- Lundqvist C, Siösteen A, Blomstrand C, Lind B, Sullivan M. Spinal cord injuries: clinical, functional and emotional status. *Spine* 1991;16(1):78–83.
- Denkers MR, Biagi HL, O'Brien AM, Jadad AR, Gauld ME. Dorsal root entry zone lesioning used to treat central neuropathic pain in patients with traumatic spinal cord injury: a systematic review. *Spine* 2002;27(7):E177–84.
- Gault D, Morel-Fatio M, Albert T, Fattal C. Chronic neuropathic pain of spinal cord injury: what is the effectiveness of psychocomportemental management? *Ann Phys Med Rehabil* 2009;52(2):167–72.
- Fattal C, Kong-A-Siou D, Gilbert C, Ventura M, Albert T. What is the efficacy of physical therapeutics for treating neuropathic pain in spinal cord injury patients? *Ann Phys Med Rehabil* 2009;52(2):149–66.
- Attal N, Mazaltarine G, Perrouin-Verbe B, Albert T. Chronic neuropathic pain management in spinal cord injury patients. What is the efficacy of pharmacological treatments with a general mode of administration? (Oral, transdermal, intravenous). *Ann Phys Med Rehabil* 2009;52(2):124–41.
- Levendoglu F, Oğün CO, Ozerbil O, Oğün TC, Ugurlu H. Gabapentin is a first line drug for the treatment of neuropathic pain in spinal cord injury. *Spine* 2004;29(7):743–51.
- Putzke JD, Richards JS, Kezar L, Hicken B, Ness T. Long-term use of gabapentin for treatment of pain following traumatic spinal cord injury. *Clin J Pain* 2002;18(2):116–21.
- Cardenas DD, Warme CA, Turner JA, Marshall H, Brooke MM, Loeser JD. Efficacy of amitriptyline for relief of pain in spinal cord injury: results of a randomized controlled trial. *Pain* 2002;96(3):365–73.
- Rintala DH, Holmes SA, Courtade D, Fiess RN, Tastard LV, Loubser PG. Comparison of the effectiveness of amitriptyline and gabapentin on chronic neuropathic pain in persons with spinal cord injury. *Arch Phys Med Rehabil* 2007;88(12):1547–60.
- Widerström-Noga EG, Turk DC. Types and effectiveness of treatments used by people with chronic pain associated with spinal cord injuries: influence of pain and psychosocial characteristics. *Spinal Cord* 2003;41(11):600–9.
- Warme CA, Turner JA, Marshall HM, Cardenas DD. Treatments for chronic pain associated with spinal cord injuries: many are tried, few are helpful. *Clin J Pain* 2002;18(3):154–63.
- Henwood P, Ellis JA. Chronic neuropathic pain in spinal cord injury: the patient's perspective. *Pain Res Manag* 2004;9(1):39–45.
- Bryce TN, Ragnarsson KT. Pain after spinal cord injury. *Phys Med Rehabil Clin N Am* 2000;11(1):157–68.
- Widerström-Noga E, Biering-Sørensen F, Bryce T, Cardenas DD, Finnerup NB, Jensen MP, et al. The international spinal cord injury pain basic data set. *Spinal Cord* 2008;46(12):818–23.
- Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113(1–2):9–19.
- Jensen MP, Barber J, Romano JM, Hanley MA, Raichle KA, Molton IR, et al. Effects of self-hypnosis training and EMG bio-feedback relaxation training on chronic pain in persons with spinal cord injury. *Int J Clin Exp Hypn* 2009;57(3):239–68.
- Amr YM. Multi-day low dose ketamine infusion as adjuvant to oral gabapentin in spinal cord injury related chronic pain: a prospective, randomized, double blind trial. *Pain Physician* 2010;13(3):245–9.
- Vranken JH, Hollmann MW, van der Vegt MH, Kruis MR, Heesen M, Vos K, et al. Duloxetine in patients with central neuropathic pain caused by spinal cord injury or stroke: a randomized, double-blind, placebo-controlled trial. *Pain* 2011;152(2):267–73.
- Heutink M, Post MW, Bongers-Janssen HM, Dijkstra CA, Snoek GJ, Spijkerman DC, et al. The CONECSE trial: results of a randomized controlled trial of a multidisciplinary cognitive behavioral program for coping with chronic neuropathic pain after spinal cord injury. *Pain* 2012;153(1):120–8.
- Salinas FA, Lugo LH, García HI. Efficacy of early treatment with carbamazepine in prevention of neuropathic pain in patients with spinal cord injury. *Am J Phys Med Rehabil* 2012;91(12):1020–7.
- Tan G, Rintala DH, Jensen MP, Richards JS, Holmes SA, Parachuri R, et al. Efficacy of cranial electrotherapy stimulation for neuropathic pain following spinal cord injury: a multi-site randomized controlled trial with a secondary 6-month open-label phase. *J Spinal Cord Med* 2011;34(3):285–96.
- Norrbrink C, Lundeberg T. Acupuncture and massage therapy for neuropathic pain following spinal cord injury: an exploratory study. *Acupunct Med* 2011;29(2):108–15.
- Arienti C, Daccò S, Piccolo I, Redaelli T. Osteopathic manipulative treatment is effective on pain control associated to spinal cord injury. *Spinal Cord* 2011;49(4):515–9.
- Kang BS, Shin HI, Bang MS. Effect of repetitive transcranial magnetic stimulation over the hand motor cortical area on central pain after spinal cord injury. *Arch Phys Med Rehabil* 2009;90(10):1766–71.
- Norrbrink C. Transcutaneous electrical nerve stimulation for treatment of spinal cord injury neuropathic pain. *J Rehabil Res Dev* 2009;46(1):85–93.
- Norrbrink C, Lundeberg T. Tramadol in neuropathic pain after spinal cord injury: a randomized, double-blind, placebo-controlled trial. *Clin J Pain* 2009;25(3):177–84.
- Rintala DH, Fiess RN, Tan G, Holmes SA, Bruel BM. Effect of dronabinol on central neuropathic pain after spinal cord injury: a pilot study. *Am J Phys Med Rehabil* 2010;89(10):840–8.
- Finnerup NB, Grydehøj J, Bing J, Johannesen IL, Biering-Sørensen F, Sindrup SH, et al. Levetiracetam in spinal cord injury pain: a randomized controlled trial. *Spinal Cord* 2009;47(12):861–7.
- Soler MD, Kumru H, Pelayo R, Vidal J, Tormos JM, Fregni F, et al. Effectiveness of transcranial direct current stimulation and visual illusion on neuropathic pain in spinal cord injury. *Brain* 2010;133(9):2565–77.
- Richardson EJ, Ness TJ, Redden DT, Stewart CC, Richards JS. Effects of nicotine on spinal cord injury pain vary among subtypes

- of pain and smoking status: results from a randomized, controlled experiment. *J Pain* 2012;13(12):1206–14.
- 39 Norrbrink C, Lindberg T, Wahman K, Bjerkefors A. Effects of an exercise programme on musculoskeletal and neuropathic pain after spinal cord injury – results from a seated double-poling ergometer study. *Spinal Cord* 2012;50(6):457–61.
 - 40 Murray DM. Design and analysis of group randomized trials. New York: Oxford University Press; 1998.
 - 41 Fisher RA. Statistical methods for research workers. 12th ed. Edinburgh: Oliver and Boyd; 1954.
 - 42 Donner A, Klar N. Design and analysis of cluster randomization trials in health research. New York: Oxford University Press; 2000.
 - 43 Ridout MS, Demétrio CG, Firth D. Estimating intraclass correlation for binary data. *Biometrics* 1999;55(1):137–48.
 - 44 Raudenbush SW, Bryk AS. Hierarchical linear models: application and data analysis methods. 2nd ed. London: Sage; 2002.
 - 45 Brown H, Prescott R. Applied mixed models in medicine. 2nd ed. West Sussex: Wiley; 2006.
 - 46 Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73(1):13–22.
 - 47 Bieler GS, Williams RL. Cluster sampling techniques in quantal response teratology and developmental toxicity studies. *Biometrics* 1995;51(2):764–76.
 - 48 Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, *et al.* Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2003;106(3):337–45.
 - 49 Melzack R. The short-form McGill pain questionnaire. *Pain* 1986;30(2):191–7.
 - 50 Richards JS, Hicken BL, Putzke JD, Ness T, Kezar L. Reliability characteristics of the Donovan spinal cord injury pain classification system. *Arch Phys Med Rehabil* 2002;83(9):1290–4.
 - 51 Putzke JD, Richards JS, Hicken BL, Ness TJ, Kezar L, DeVivo M. Pain classification following spinal cord injury: the utility of verbal descriptors. *Spinal Cord* 2002;40(3):118–27.
 - 52 Baastrup C, Finnerup NB. Pharmacological management of neuropathic pain following spinal cord injury. *CNS Drugs* 2008;22(6):455–75.
 - 53 Allison PD. Fixed effects regression methods for longitudinal data using SAS. Cary, NC: SAS Institute; 2005.