

EXPERT REPORT OF NICHOLAS P. JEWELL, Ph.D.

Executive Summary of Report

1. Backonja et al.¹ report on the clinical efficacy of gabapentin on subjective pain scores as compared to placebo, based on a randomized controlled clinical trial of patients with diabetes mellitus suffering from painful diabetic neuropathy. These authors consider a potential bias in their results due to the significantly higher incidence of side effects among subjects exposed to gabapentin that might “unblind” an individual’s treatment assignment, thereby distorting their subsequent pain reports. However, the authors did not find evidence of such a bias in a very simplistic analysis of two side effects analyzed separately. I have reanalyzed their data more thoroughly. The summary of my findings is that almost the entire apparent treatment effect reported in Backonja et al.¹ disappears when data after the occurrence of treatment-related Central Nervous System (CNS) side effects is eliminated. I conclude that the trial provides no basis of any clinical efficacy for gabapentin over placebo in reducing pain in this population.

Qualifications

2. For the past 27 years, I have been a Professor in the Division of Biostatistics, School of Public Health, and in the Department of Statistics, both at the University of California, Berkeley. Specifically, I have served as a Full Professor (1987 – present); Associate Professor (1982 – 1987); and Assistant Professor (1981 – 1983). Prior to that, I was an Assistant Professor in the Department of Statistics at Princeton University, Princeton, New Jersey (1979 – 1981) where I also served as Director of the Statistical Laboratories. At Berkeley, I also held the position of Chair of the University

¹ “Gabapentin for the symptomatic treatment of painful neuropathy in patients with diabetes mellitus. A Randomized Controlled Trial,” Backonja, M., Beydoun, A., Edwards, K. R. et al., JAMA, 1998, 280, 1831-1836.

of California Graduate Group in Biostatistics from 1986 – 1994, and from 2000 – 2007. From 1994 to 2000, I served as Vice Provost at the University of California, Berkeley, in the Office of the Chancellor. Since September 2007, I have been serving as Vice Provost, Academic Personnel, at the Office of the President of the University of California. A true copy of my CV is attached as Exhibit A.

3. I have served as a member of the National Academy of Sciences Committee on National Statistics (1993 – 1996), and of the Committee on Theoretical and Applied Statistics (1994 – 1996). I received my Ph.D. in Mathematics from the University of Edinburgh, Scotland in June 1976, and was a post-doctoral Harkness Fellow at Stanford University and the University of California at Berkeley from 1976 to 1978. In 1978-79 I was a Research Fellow in the Medical Statistics Unit at the University of Edinburgh, Scotland. I also held a Visiting Professor appointment at Oxford University, England, in Spring 1990, and at the London School of Hygiene and Tropical Medicine in Spring 2007. During April – May 2007, I was a resident Fellow of the Rockefeller Foundation at their Study Center in Bellagio, Italy.

4. I am the author of a textbook, *Statistics for Epidemiology*, (Chapman and Hall, New York 2003), as well as approximately 120 peer-reviewed articles in the field of biostatistics. My fields of expertise include the analysis and interpretation of survival data and other statistical methods to investigate risk factors for disease outcomes, and longitudinal data analysis. I am a founding editor of the *International Journal of Biostatistics*, Senior Editor for the journal *Statistical Applications in Genetics and Molecular Biology*, and Associate Editor for the journal *Biometrika*. In 2005, I received the Snedecor Award, from the Committee of Presidents of the Statistical Societies, awarded to “an individual who was instrumental in the development of statistical theory in biometry.” The award is associated with the best publication in biostatistics in the world in the previous 3 years. I also received a Distinguished Teaching Award from the School of Public Health, University of

California at Berkeley, in 2004.

5. I am or have been a member of several international statistical societies including the International Biometric Society, the Institute of Mathematical Statistics, and the American Statistical Association. I was made a Fellow of the American Statistical Association in 1991, and a Fellow of the Institute of Mathematical Statistics in 1996. I served as President of the Western North American Region of the International Biometric Society in 1991 – 92, and as Treasurer of the Institute of Mathematical Statistics from 1985 – 1988. In 2007, I was made a Fellow of the American Association for the Advancement of Science (AAAS).

6. In the past 5 years, I have testified at deposition in a Blood Products case, a case concerning the malfunctioning of medical devices, and twice on a case involving the adverse cardiovascular effects of celebrex. To the best of my knowledge, the relevant cases numbers are as follows:

In re: Factor VIII or IX Concentrate Blood Products Litigation, MDL NO. 986, N.D. Ill. Case No. 1:93CV7452

In re: Guidant Implantable Defibrillators Products Liability Litigation, MDL No. 1708, D. Minn. Case Nos. 05-1708, 06-00025 and 05-02596

In re: Bextra and Celebrex Marketing Sales Practices, and Product Liability Litigation, MDL NO. 1699, N.D. Cal. Case No. M:05-CV-01699 CRB

My consulting rate for this project is \$350 per hour. I have relied on the documents referenced in this report, and upon my experience in the field of biostatistics, in preparing this document.

Gabapentin for Treatment of Painful Neuropathy in Patients with Diabetes

7. I have been asked to provide an analysis of data underlying the paper by Backonja et al.¹ on the effects of gabapentin on pain severity in a sample of patients randomized to treatment with either gabapentin or placebo. I was supplied with

the original data in electronic form, including daily pain reports for each subject during a short screening period, and then during the eight-week trial. Subjects were randomized to receive either gabapentin or placebo, and the randomization indicator was also provided in the data sources. Patients were recruited from 20 clinical sites. The randomization was carried out in a double-blind fashion. The primary efficacy outcome was a pain severity rating, recorded by each patient in daily diaries using an 11-point Likert scale ranging from 0 (“no pain”) to 10 (“worst possible pain”). The study, including dosing procedures, is described in further detail in Backonja et al.¹. The data provided information on 81 placebo subjects and 84 patients receiving gabapentin, as reported in Backonja et al.¹

The double blinding of subjects is potentially compromised by several factors, one being the occurrence of side effects known to be related to the study medication, gabapentin. This is of particular concern given that the primary outcome is only measured subjectively. Backonja et al.¹ raise this issue directly in stating “Because the study end point of pain was subjective, we explored the possibility that the occurrence of adverse events resulted in the “unblinding” of the study, biasing the results of our efficacy analysis.” I confirm that this form of “unblinding” is well known, to be expected in pain studies, and has been studied extensively.² The authors noted that “Dizziness and somnolence, the 2 most frequent adverse events, were also those with the largest difference in incidence between the gabapentin and placebo groups,” considerably raising the possibility of bias. After an ad hoc approach that involved ignoring all data from those patients who experienced a single adverse effect (such as dizziness), the authors concluded “inclusion of patients who experienced these central nervous systems adverse effects in the original analysis did not account for the overall efficacy seen in the trial.”

I was asked to review the assessment and estimates of treatment efficacy, paying particular attention to whether such results were influenced, or biased, by the occurrence

² “Blinding effectiveness and association of pretreatment expectations with pain improvement in a double-blind randomized controlled trial,” Turner, J. A., Jensen, M. P., Warms, C. A. et al., *Pain*, 2002, 99, 91-99.

of side effects considered associated with the study medication.

Basic Analysis of Pain Outcome Data Ignoring Occurrence of Side Effects

8. I first fit a simple longitudinal regression model using the data on pain scores for each day to understand the trajectory of pain assessments over time, and how changes differ for patients in the two treatment groups. This is the most direct way to approach the data, avoiding compressing information on the pain outcome into weekly averages, and, in particular, into an endpoint measured “as the mean score for the last 7 diary entries.” The specific regression model assumes that the mean pain score varies linearly over time. Estimates of the regression coefficients were obtained using the standard generalized estimating equation (GEE) approach for linear regression, thereby allowing for correlation amongst repeated pain scores on the same individual³. The software used was Stata 9.2⁴.

9. The simplest regression model used assumed a linear relationship of the average pain score over time, common to all patients, but allowing a different intercept and slope for the two treatment groups. At this point, all follow-up pain data is being exploited, ignoring information on side effects that might “unblind” treatment assignment. For the placebo group, the estimated intercept (interpretable as the mean pain score at day 0 as the study began) was 6.11 with an estimated subsequent daily decline of 0.027 per day (or, equivalently, 1.52 over 56 days). This presumably represents some placebo effect or simply systematic improvement in patients’ conditions after recruitment. On the other hand, the estimated intercept for the active treatment group is 5.61 with an estimated daily decline of 0.045 per day (or, equivalently, 2.52 over 56 days). The rate of decline is significantly greater for the active treatment group ($p = 0.005$). It is somewhat surprising that the intercepts are also significantly different but the

³ *Analysis of Longitudinal Data*, Diggle, P., Heagerty, P., Liang, K-Y., Zeger, S., 2nd Edition, 2002, Oxford University Press.

⁴ Stata/SE for Macintosh, Stata Corporation, College Station, Texas.

p-value of 0.045 is borderline.

10. These calculations used the screening measurements of pain before treatment began. Removing these observations and using only pain scores from day zero yields very similar results. The estimated intercepts and slope for the placebo group are now 6.01 and -0.025 per day (a decline of 1.38 over 56 days). For the active treatment group, the analogous estimates are 5.29 and -0.037 per day (a decline of 2.06 over 56 days). Now the slopes are not significantly different ($p = 0.069$) although the intercepts appear slightly more different ($p = 0.016$).

11. The above analysis has assumed that subjects recruited at different clinical sites report similar pain scores. To allow for the possibility that patients from different sites experience different levels of pain, we introduced site as a covariate exactly as specified in the Backonja et al.¹ analysis. Adjusting for clinical site in this way allows the regression line, describing changes in mean pain scores over time, to have a different intercept for each site, but assumes that the rate of decline in pain scores is the same for all sites for both the placebo and active treatment groups. The adjustment for site changes the estimated rate of decline for the two groups only slightly to 0.024 per day (a decline of 1.34 over 56 days) for placebo patients, and 0.037 per day (a decline of 2.10 over 50 days) for the active treatment group. Comparison of these rates of decline yields a p-value of 0.050. As noted, in this analysis, the intercepts are allowed to vary from site to site, but overall the average differences in intercepts between placebo and active treatment groups (for a given site) is estimated to be 0.61 lower in the gabapentin group, remaining just statistically significant ($p = 0.032$).

12. In summary, there is marginal evidence for both an immediate drop in pain scores in the treatment group followed by a slightly greater decline in pain over the course of treatment. Adding the differences after an eight week (56 days) follow-up period produces an estimated average pain score drop of around 1.4 more on active treatment than on placebo.

Adjustment for Baseline Pain Scores

13. Backonja et al.¹ also used the screening mean pain score, measured over a brief period prior to commencement of treatment as an additional covariate. Due to randomization there should be no difference in the screening mean pain score across the two treatment groups (and this is supported by the data), and so inclusion of this covariate should not alter the previous estimates discussed above. Nevertheless, there is some possibility of additional precision since it is likely that the screening mean pain score is closely correlated with subsequent pain score measurements (again confirmed by the data). To follow Backonja et al.¹ therefore, we extended the above analysis to control for where a patient “starts” in terms of their screening mean pain score. Holding site fixed as before, the above estimates change as follows: the rate of decline for placebo patients is now 0.027 per day (a decline of 1.49 over 56 days), and 0.037 per day (a decline of 2.06 over 56 days) for the active treatment group. These rates of decline differences have an associated p-value of 0.14, no longer statistically significant. Overall the average differences in intercepts between placebo and active treatment groups (for a given site) is estimated to be 0.59 lower in the gabapentin group, remaining statistically significantly different from zero ($p = 0.006$). As noted, there is a very strong association between pain scores after treatment and the screening mean pain score, and this relationship is similar for the two treatment groups. These results are not changed dramatically by including (non-significant) interaction terms that allow the rates of decline to depend on the average pain score at screening.

Summary of Basic Analysis

14. There is evidence for both an immediate drop in pain scores in the treatment group followed by a slightly greater decline in pain over the course of treatment. Adding these two effects produces an estimated average pain score drop of around 1.2 more on active treatment than on placebo, after an eight-week (56 days) follow-up period. As these estimates control for the effects of site and screening mean

pain score, this analysis closely corresponds with that of Backonja et al.¹ who report an overall increased drop of 1.2 in the gabapentin group as compared to the placebo subjects. These authors provide an associated overall p-value less than 0.001 for this comparison. In my analysis, the equivalent p-value is also less than 0.001. Thus, to this point, my analysis agrees closely with Backonja et al.¹

I now turn to consideration of the effects of the occurrence of adverse side effects that are considered related to treatment and that therefore have the potential to “unblind” treatment assignment and subsequently bias the results.

Stopping Analysis at Onset of Treatment-Related Side Effects

15. To address the issue surrounding the occurrence of side effects I examined the data to determine the first day on which a treatment-related side effect occurred. To judge “relatedness”, I used a variable denoted “aerel” and considered the side effect related to study medication if aerel was assigned a value in the range 3—5 (indicating that the side effect was possibly, probably, or definitely related to the study medication). Finally, I used only the first Central Nervous System (CNS) side effect as defined by a list in Appendix C.36 of the 945-210 research report, or a synonym of any condition found there (for example, sleepiness included with somnolence). For the placebo group, 12 individuals experienced such a related side effect during their period of follow-up; on the other hand, 44 members of the gabapentin group had such side effects in the follow-up period. These numbers reflect what Backonja et al.¹ indicated regarding the higher incidence of treatment related side effects in the gabapentin treated subjects. For the 12 members of the placebo group who experience treatment related side effects as described, the average day of onset is 12.1 days with a standard deviation of 8.6 days; for the 44 analogous subjects in the gabapentin group, the average day of onset is 11.0 days with a standard deviation of 9.2 days.

In the placebo group, the nature of these observed “first” related side effects is distributed across an entire range from dizziness to tiredness, with somnolence (plus its

synonyms) experienced by seven subjects. Another three subjects suffered from dizziness (including synonyms). For the gabapentin group, there is again a wide variety of first side effects experienced from ataxia to vertigo. In this group eighteen subjects suffered dizziness (including synonyms) as the first treatment related side effect, thirteen somnolence (including synonyms), and four confusion. No other included first side effect was experienced by more than one subject in either group.

To assess the possible impact of these treatment related side effects “unblinding” treatment assignment and subsequently biasing the subjective measurement of pain, I removed all data following the first occurrence of a treatment related CNS side effect as defined above. This is a less radical approach than in Backonja et al.¹ who removed the entire record for subjects who reported such side effects, albeit only for two major side effects and only one side effect at a time.

There is a possibility that the results of paragraphs 9-11 (and the conclusion of paragraph 12) will be modified if we only use the data up to and including the first day of onset of a side effect related to treatment, but no pain scores thereafter. In particular, I would anticipate a slight increase in variability due to using less data. However, our primary interest is to see if the estimates of treatment effects themselves are fundamentally changed, since this would likely be due to bias in pain measurements after the occurrence of these side effects. The following paragraphs therefore focus on this issue.

16. Repeating the most general analysis (without adjusting for average pain at baseline) reveals a somewhat different picture from the results in Paragraph 11. Now (adjusting for site), the estimated rate of decline is 0.025 per day (a decline of 1.38 over 56 days) for placebo patients, with *essentially the same effect* for the active treatment group with an estimated decline of 0.020 per day (a decline of 1.12 over 56 days). Note that the rate of decline is actually slightly slower in the active treatment group: the comparison of the rates of decline between placebo and active treatment is

now entirely insignificant ($p = 0.49$), indicating no apparent treatment difference. Thus, the apparent difference in the rates of decline between the two groups is completely explained by pain scores that occur after treatment-related side effects have occurred.

On the other hand, in this analysis, the overall average difference in intercepts between placebo and active treatment groups (for a given site) is 0.67 (with the smaller pain scores in the treatment group), basically unchanged from what I reported in paragraph 11, a difference that remains just statistically significant ($p = 0.02$).

17. Continuing in this way, we now repeat the analysis that adjusted for the average screening pain score described in paragraph 13. Holding site fixed, the estimates change as follows: the estimated rate of decline for placebo patients is 0.029 per day (a decline of 1.60 over 56 days), and 0.021 per day (a decline of 1.18 over 56 days) for the active treatment group, again more decline observed with placebo than under active treatment. These rates of decline differences are again essentially equivalent with an associated p -value of 0.42. Overall the estimate average difference in intercepts between placebo and active treatment groups (for a given site) is 0.57 (lower average pain scores under active treatment), again just statistically significant ($p = 0.013$).

18. Adding the two treatment effects, described in paragraph 17, produces an estimated average pain score drop of just 0.15 more on active treatment than on placebo, after an eight-week (56 days) follow-up period. The smaller number reflects the now increased decline under placebo that essentially offsets the initial overall difference due to treatment. Effectively there is no difference between pain scores under placebo and those under gabapentin after eight weeks of treatment.

19. Illustrating the data for a single patient powerfully displays the bias that can be introduced by including data after an adverse side effect first appears, possibly alerting the patient that they may be on active treatment. Figure 1 below graphs the pain score data by day for subject #4017 who was randomized to the gabapentin group, and first reported the side effect of somnolence on day 11 of treatment.

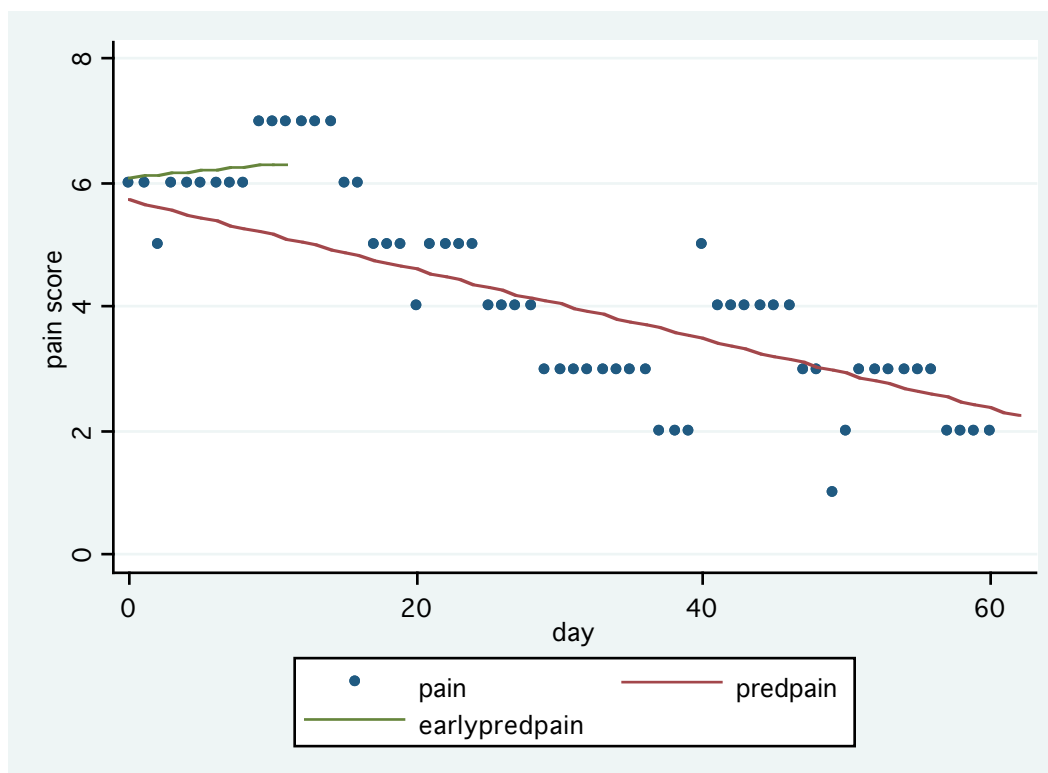


Figure 1. Plot of Pain Scores by Day for Subject #4017, with Regression Lines Using (i) All Data (long red line), and (ii) Only Data for Days up to and Including the Day of the First Observation of a Treatment Related Side Effect, namely Day 11 (short black line).

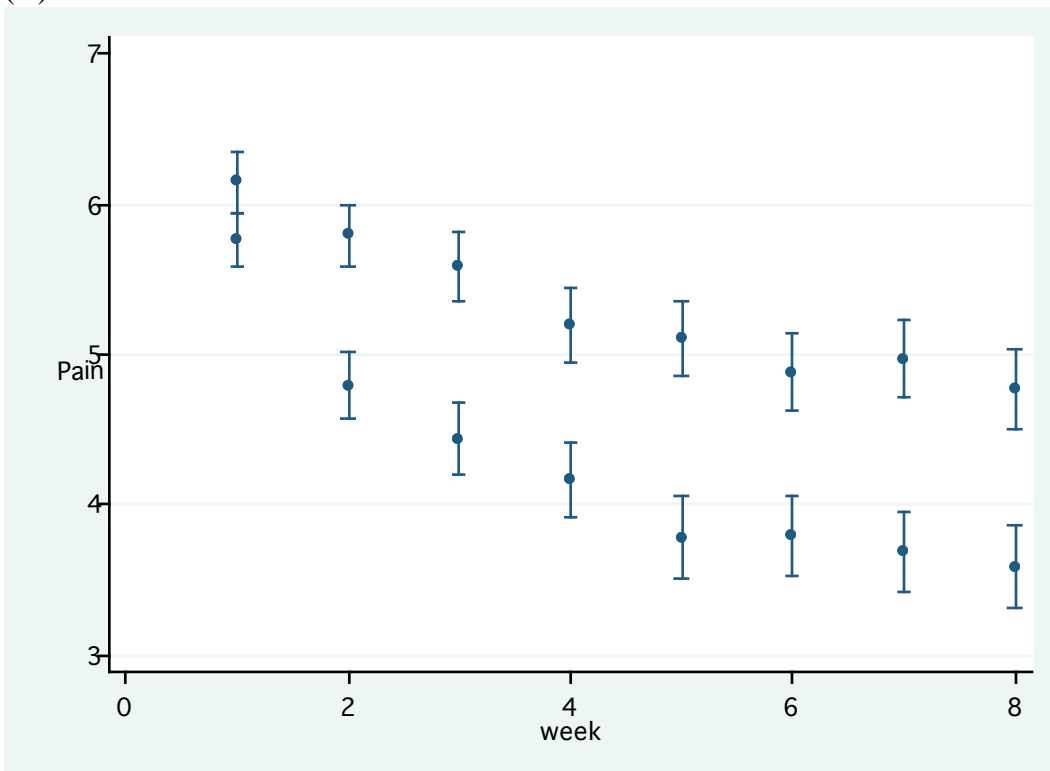
Note that, for this individual, pain scores were generally *increasing* over the first 11 days of treatment. Within a few days of the first noted occurrence of a treatment related side effect, the subjective pain scores systematically decline. The long red line in Figure 1 models the change in mean pain scores over time over the entire period, and suggests a systematic decline in pain over time. However, the analysis restricted to the first 11 days--before the possible “unblinding” of treatment assignment--paints a very different picture.

Of course, Figure 1 displays the results for only a single subject whose side effect occurs within the second week of treatment; however, the analyses described in paragraphs 16—17 effectively “average” these estimated regression lines over all patients, if and when treatment-related side effects occur, allowing for a different

intercept from subject to subject depending on their mean screening pain score and their clinical site. The latter analysis shows that the phenomenon displayed in Figure 1 is not unusual but systematic over the sample.

20. Finally, we present further graphic evidence and analyses supporting the results previously described. Specifically, we display estimates of weekly mean pain scores in both the gabapentin and placebo groups, before and after adjusting for the occurrence of potentially “unblinding” CNS side effects. Specifically, Figure 2(A) below mimics Figure 2(A) in Backonja et al.¹ (p. 1834). However, Figure 2(B) shows far less difference between the gabapentin and placebo groups after accounting for the possibility of “unblinding” due to the occurrence of CNS treatment related side effects. In Figure 2(A), a simple comparison of the weekly means between the placebo and gabapentin groups is statistically significant (p -value < 0.05) from week 2 through week 8 inclusively, as described in Backonja et al.¹ However, in Figure 2(B), the means are only statistically significant in weeks 2 and 3, and never thereafter. In particular, at week 8, the p -value for the comparison of mean pain scores between the gabapentin and placebo groups is 0.41, suggesting no difference between the two groups. When this comparison is adjusted for site and screening mean pain, the p -value for treatment efficacy is 0.81, with an estimated drop in mean pain score due to treatment of -0.09 with an associated 95% confidence interval of $(-0.84, +0.65)$. This is the analysis that most closely corresponds to that reported in Backonja et al.,¹ and indicates that it is just as likely that placebo reduces the pain score at the end of eight weeks as gabapentin. The small estimated difference--as shown, for example, in Figure 2(B) at week 8, before adjustment for covariates--is likely due to bias as previously discussed.

(A)



(B)

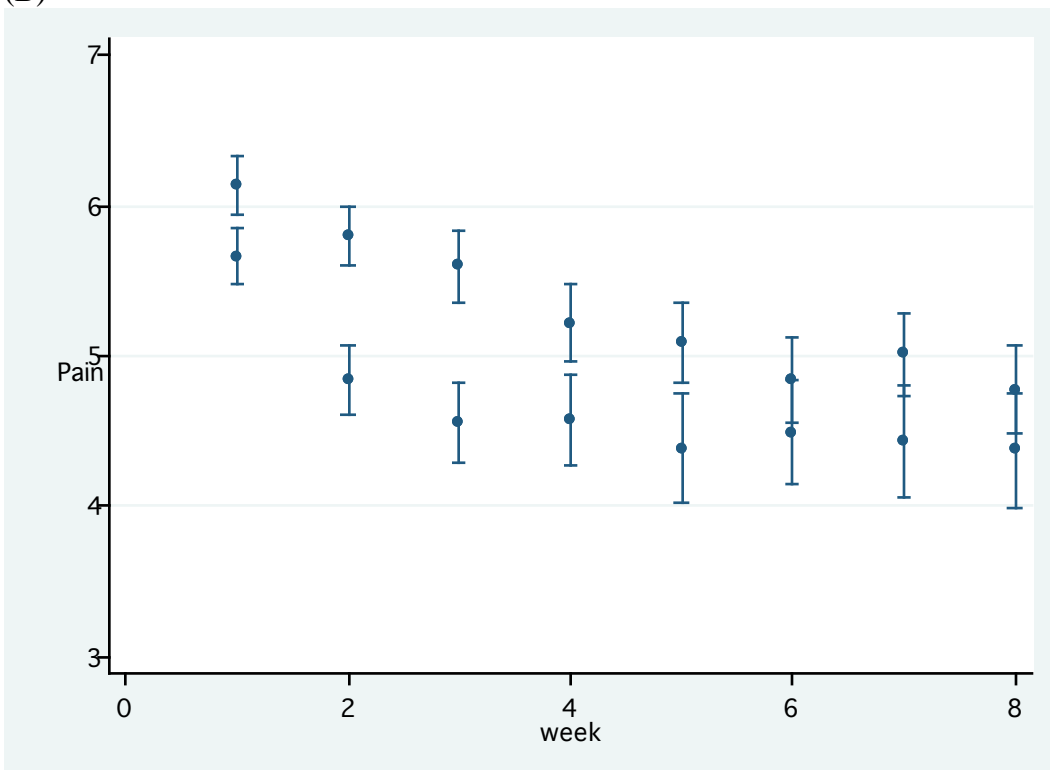


Figure 2. Weekly estimated mean pain scores by treatment group (placebo: upper dots; gabapentin: lower dots) with one standard error bars. (A) Using all data; (B) Removing data after occurrence of CNS treatment-related side effect.

Conclusion

20. Approximately 90% of the apparent total improvement in average pain scores under active treatment reported by Backonja et al.¹ is explained by events after the onset of adverse side effects that have the potential to “unblind” treatment assignment, and are therefore subject to bias. The estimated remaining difference between placebo and gabapentin, of 0.15 (as described in paragraph 18) or 0.09 (as described in paragraph 20), is small and difficult to interpret since it is based on an effect that is “present” at the very onset of treatment and not improved by further treatment (in fact, worsened). In the absence of any difference in decline once treatment has begun, it is hard to accept that this is due to a true effect of treatment, and it may be better understood as the impact of some form of additional unanticipated bias. For example, given the considerable and demonstrated impact of observed treatment related side effects, presumably due to “unblinding,” there is a similar potential for further bias caused by more subtle—unreported—experience of side effects and other factors associated with treatment, that may also have contributed to “unblinding” thereby exhibiting the small treatment-placebo “difference”. Given that a simple adjustment of the data to allow for reported CNS side effects, related to treatment, eradicates almost all treatment effect, it is not unlikely that the small remaining difference could be explained by small differences in unreported side effects that also might “unblind” treatment assignment. My findings entirely contradict the claim by Backonja et al.¹ that “inclusion of patients who experienced these central nervous systems adverse effects in the original analysis did not account for the overall efficacy seen in the trial.” It is my view that my new, more thorough, analysis completely undermines the claims of treatment efficacy made in Backonja et al.¹

21. I reserve the right to supplement this report if new or significantly modified quantitative information is provided at any point, or I have the opportunity to review further reports or published literature.



July 29, 2008

Nicholas P. Jewell, Ph.D.

Bibliography

Backonja, M., Beydoun, A., Edwards, K. R. et al., "Gabapentin for the symptomatic treatment of painful neuropathy in patients with diabetes mellitus. A Randomized Controlled Trial," *JAMA*, 1998, 280, 1831-1836.

Diggle, P., Heagerty, P., Liang, K-Y., Zeger, S., *Analysis of Longitudinal Data*, 2nd Edition, 2002, Oxford University Press.

Turner, J. A., Jensen, M. P., Warm, C. A. et al., "Blinding effectiveness and association of pretreatment expectations with pain improvement in a double-blind randomized controlled trial," *Pain*, 2002, 99, 91-99.