Dear Dr. Burch and Members of the Editorial Board,

Thank you for your meticulous review of our revision. We would like to reinforce that we are absolutely committed to seeing this through to fruition at BMJ. We feel that all of your comments are (and were) reasonable, and that there are not irreconcilable differences. Rather, we seemed to have misunderstood some of the intent of the previous comments in that we (mistakenly) thought that in some places the Editorial Board wanted clarifications or explanations (and left that at our discretion), rather than changes. We have therefore made every single change that you requested, adding multiple new references (while trying to keep the manuscript at a reasonable length.). We hope you more satisfied with this version.

Very respectfully,

Steven Cohen & Co-Authors

Many of our statistician's comments have not been addressed fully. Most importantly, we require trial reporting to adhere exactly to the trial registration, not the protocol. If there are discrepancies between the registration and the protocol, we ask that authors adhere to the trial registration. The primary outcomes described in the trial registration should be presented first. If amendments were made to the trial registration or trial protocol, it should be made very clear in the methods what changes were made, the specific reasons why these changes were made, and when they were made. Any analyses made as a result of amendments to the trial registration or protocol can be presented in the paper, but it should be clear that these are secondary analyses. If any changes possibly affected the outcome of the study, this should also be made clear.

Answer: We have done as you asked in the abstract and the beginning of the results on page 13. The trial registration lists leg pain at 1 and 3 months as primary outcome measures and the protocol specifically states, "The primary outcome variable will be the 0-10 numerical rating scale leg pain score" (i.e. there is no discrepancy).

*On the question of terminology regarding back and/or leg pain, thank you for the extensive discussion in your response. It's an interesting and helpful description of how current thinking around causes of back and leg pain may be in evolution. In reporting results of a trial, however, the terminology used in the paper should reflect the intention at the beginning of the study. In looking through the study protocols and trial registration I note that the term used in the study objectives and procedures was "lumbosacral radiculopathy". I also note that the inclusion criteria in the protocol include the term "lumbosacral radicular pain". Unless there were changes to either the study objectives or the inclusion criteria during the study, it seems most reasonable to reflect the terminology used when the study was conceived. Thus where the term "sciatica" is used, "lumbosacral radiculopathy" should be used instead. I do agree that in some cases, the specialty agrees on a clear change of terminology between the time when the study starts and when it is reported, but that doesn't seem to be the case here.

Answer: We have changed it to lumbosacral radicular pain as suggested since someone with "radiculopathy" does not necessarily have to have pain, and we measured leg pain (though we would be willing to change it to radiculopathy as well since all of these patients had radiculopathy).

It might also be helpful to include a sentence or two describing why both back and leg pain are recorded, similar to what you described here: "There may be cases of 'axial-only' back pain that are neuropathic in nature, and of course not all leg pain is neuropathic…Because patients cannot accurately distinguish between mechanical and neuropathic pain scores, leg (indicative of neuropathic pain) and back pain scores are always used instead."

Answer: Under the section, "Co-Interventions, Outcome Measures, Follow-up & Missing Data" on page 10, we have added the following statement, with 4 references to support it: "Because patients cannot accurately distinguish between neuropathic pain (which is more likely to respond to ESI and gabapentin) and mechanical pain, leg (neuropathic) and back (mechanical) pain scores were separately recorded as surrogates. Whereas a small percentage of cases of axial back pain may be neuropathic in nature, and many forms of mechanical pain such as facet arthropathy and degenerative disc disease may radiate into the leg, validated instruments have shown that leg pain emanating from the back is generally indicative of radicular pain."

*Inclusion and exclusion criteria: How was radicular leg pain specifically defined? Please specify a comprehensive list of signs and symptoms that allowed inclusion into the trial. Please specify a list of medical or psychiatric conditions considered part of the exclusion criteria. The methods should be described in enough detail that an independent researcher would be able replicate the study based on what's written. The inclusion and exclusion criteria are detailed quite well in the study protocol submitted to Walter Reed, so just including all of the information there might be helpful. (I note further down in the response that this was "left up to the provider". Is it possible that different providers would have included or excluded a different set of patients, and can the authors comment on whether this introduces heterogeneity?)

As noted above, on page 10 we noted as requested why "leg pain" was considered radicular, and why we separately recorded back pain. We have elaborated on page 6 the list of inclusion and exclusion criteria, and included the cutoffs we used for depression and PTSD, with references (note, only if a patient reported or had a history of depression was he or she administered the BDI). At the time that the First Author is writing this, he is on a plane to Montreal to give a talk at McGill on "Clinical Trial Design". We believe that there is always going to be heterogeneity in clinical trials, and it is often written that heterogeneity is desirable in comparative-effectiveness research to better simulate "real life" conditions. However, to enhance selection consistency (really eliminate errors and reduce bias), we performed briefings and debriefings (to align communication and methods at both military and academic centers), ensured everyone had access to a Senior Investigator to answer questions, and performed practice enrollments at one site (Walter Reed). This is noted at the end of the "Participants and Settings" section of the "Methods". More recently (for our next clinical trial), we have started to require all investigators to view slide presentations as well.

The wording around the imaging criteria is still not clear. Would something like "All included patients were required to have MRI findings of herniated disc or spinal stenosis" be acceptable?

On page 6, we add the following statement as requested, "All participants were also required to have MRI findings of a herniated disc or spinal stenosis concordant with their presentation."

*The question of patients being included for symptoms of neurogenic claudication vs lumbosacral radiculopathy is still somewhat unclear. This was something of a sticking point for both editors and reviewers. In both of the study protocols, the only inclusion criteria were for lumbosacral radicular pain. How were patients with neurogenic claudication entered into the study if they did not meet the inclusion criteria in the study protocols? It might be that this is a deviation from the study protocol, in which case this should be explained clearly in the methods. The subgroup analysis showing that the inclusion of these patients did not significantly affect the results is reassuring, but unexplained deviation from study protocol is concerning.

The definition of radicular pain is, "Radicular pain is a type of pain that radiates into the lower extremity directly along the course of a spinal nerve root. Radicular pain is caused by compression, inflammation and/or injury to a spinal nerve root arising from common conditions including herniated disc, foraminal stenosis and peridural fibrosis."

The definition of neurogenic claudication (from Wikipedia, which is consistent with the main textbooks in neurology and surgery) is, "Neurogenic claudication (NC) is a common symptom of lumbar spinal stenosis, or inflammation of the nerves emanating from the spinal cord." Neurogenic means that the problem originates with a problem at a nerve, and claudication, from the Latin term for limp, indicates that the patient feels a painful cramping or weakness in the legs.

As such, because the pain associated with neurogenic claudication originates from the nerve roots (nerves emanating from the spinal cord, which ends much higher) and is experienced in the legs, it should be considered as lumbar radicular pain. If one considers these definitions, then one can have symptoms related to neurogenic claudication (weakness) in the legs without radicular pain, but if there is pain- which was an inclusion criterion in our study- then it should be considered radicular pain.

In the 2nd to last sentence of "Participants and Settings" in the "Methods" on page 7, we added, "Since our intention was to evaluate the relative effects of 2 commonly used treatments for radicular pain, patients with neurogenic claudication from spinal stenosis who did not report lower leg pain (i.e. those who had only weakness or paresthesias) were excluded from participation.

Is this reasonable (we'd be open to other suggestions)?

* The subgroup analyses performed should be specified in the methods section.

The "Methods" section has been amended to include language as described in the study protocol, and the methods section now states that "Subgroup analyses were conducted to ascertain which demographic and clinical variables are associated with outcome." Additionally, specification of the subgroup analyses performed has been further detailed in the methods: "Variables examined included etiology (e.g. stenosis vs. herniated disc), level of injection, pain duration $\geq$ 3 months, injection type (i.e. transforaminal ESI for unilateral pain vs. interlaminar ESI for bilateral pain), smoking status, military status, presence of psychiatric disease, obesity, age, gender, or dose of gabapentin." The only subgroup analysis specified in the protocol that was not performed was the effect of war-related traumatic injuries (e.g. traumatic brain injury) because there were far too few people to make any analysis meaningful. These changes are noted the statistical section of the "Methods" on pages 11 and 12).

*The rationale for the target outcome of 1 point difference was not included in the study itself. Since both editors and reviewers raised this as an important question, the choice of endpoint should be very clearly supported in the text of the paper.

We have referenced 3 articles for this. The first is the IMMPACT guidelines that specifically note that although a 30% decrease should be considered clinically meaningful for a given patient, less than this can represent a clinically important effect when comparing 2 treatments in clinical trials. We also cited the studies that led to the approval of the only adjuvant (duloxetine) in the U.S. for LBP (1-point decrease in pain between duloxetine and the control) and the 2nd controlled study that led to the approval of gabapentin for PHN (also a 1-point difference between treatment groups; of note, the other study for PHN required a 1.5 point difference in pain between the gabapentin and the placebo group). The following statement was added at the very end of the "Methods" section on page 12, "Although a 30% or 2-point decrease in pain has been shown to represent a clinically meaningful benefit to an individual patient, the same IMMPACT guidelines note that smaller differences between groups can be considered to be clinically important in clinical trials. A 1-point difference in pain scores was chosen because it is consistent with U.S. Food and Drug Administration requirements for approval of adjuvant analgesic medications for LBP, and gabapentin for neuropathic pain."

*It might be appropriate to include something like your sentence about wanting to focus on comparative efficacy research in the limitations section. "The reason we did this was that we felt that there were over 50 placebo-controlled trials comparing ESI or gabapentin to placebo for sciatica, and that (per HHS recommendations) there was a strong need for "comparative-effectiveness" research."

Thank you very much for this recommendation and allowing us to do this. In the first paragraph of the discussion on page 16, we added, "In accordance with the U.S. Department of Health and Human Services recommendation to

increase and improve comparative-effectiveness research, we decided to compare 2 of the most common treatments for lumbosacral radicular pain in a double-blind fashion.", along with the reference for the actual report to the President and Congress."

*It would be helpful to specifically answer reviewer 2's question about single level ESI in the text of the paper.

At the end of the paragraph entitled "Epidural Injections" in the "Methods" on page 8, we added the following statement, along with 3 references to support it: "Whereas some physicians perform multi-level transforaminal injections, we limited our procedures to a single-level because of the increased risks this would entail, because there are no controlled studies validating the use of more than one injection, and because a well-placed transforaminal ESI generally spreads to multiple levels."

*The paper should reflect the fact that 17 year olds were allowed to join the study, as per the response here, they would have been included had any been eligible.

This is now noted in the "Methods" section on page 6 under "Participants and Settings".