PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Comparing the effects of two spasticity management strategies on the long-term outcomes of individuals with bilateral spastic cerebral palsy: a multi-center cohort study protocol
AUTHORS	Munger, Meghan; Chen, Brian Po-Jung; MacWilliams, Bruce; McMulkin, Mark; Schwartz, Michael

VERSION 1 - REVIEW

REVIEWER	Rory O'Sullivan Central Remedial Clinic, Dublin, Ireland.
REVIEW RETURNED	04-Dec-2018

GENERAL COMMENTS	This is an extremely interesting study protocol. The need for appropriately matched control groups in SDR outcome research and CP research more generally has been well highlighted in the recent literature and is well summarised in this manuscript. The study itself should add significantly to the current literature and the study protocol is certainly worthy of publication in its own right and would be of significant interest to researchers in this area as the difficulty in combining differing spasticity assessments in particular either across centres, or historically within a single centre, is common to many motion laboratories and centres. The authors have proposed a unique and novel method of dealing with combining different methods of assessing spasticity. As mentioned above many centres struggle in dealing with historical records of spasticity assessment which has changed over time and overall the proposed method looks very interesting and valid and will allow participants to be categorised as mild/moderate/severe. However, the questioned asked in the review form above is- "Are the methods described sufficiently to allow the study to be repeated?" and the description given is not quite clear enough to say 'yes' . However, I appreciate that this may well be much clearer when seeing it 'in action' in the actual study paper and is possibly part of the difficulty in writing/reviewing a protocol only. The spasticity score for GIL and SPK is based on ashworth score and is clear. The paper then states that "subjects needed to have at least 3 of 4 spasticity measures on each side to be included". This isn't quite clear and assume this just means they needed an ashworth score recorded for at least 3/4 muscles on each side to be considered for inclusion in the study rather than a particular score signifying spasticity to be present before being included? The score for SLC is less clear. Again, assume the statement that subjects need 3/5 measures on each side means they just needed

	3/5 measures recorded rather than referring to a measure signifying spasticity present? Also, the method of scaling some of these measures on a zero to five range is not clear (particularly Duncan Ely, clonus, deep tendon reflex and ankle stretch). When referring to difference between ankle initial stretch and final range does that initial stretch mean a fast stretch? Just wondering also, why weight the Ashworth scores by factor of two? Is this a reflection of its relative importance compared to the others (which I'm not disagreeing with!) or for mathematical/data management reasons? Overall, very interesting protocol of significant interest and benefit to CP researchers but bit more clarity needed on spasticity scoring proposed

REVIEWER	Nico Enslin Red Cross Children's Hospital Cape Town South Africa
REVIEW RETURNED	12-Dec-2018

GENERAL COMMENTS	Thank you for performing this valuable study. i think it may add greatly to our current understanding of the effect and place of SDR compared with other treatments. My biggest concern: I am not sure from the protocol if you will have properly compared what you think you are comparing: SDR vs non-SDR on function, etc. The authors have to bare in mind that Botox, as well as ITB during the course of the child's life will affect tone, function and therefore, outcomes also. This is therefore not a SDR vs none study. In both groups tone are influenced by treatment. It is just the method that differs. There is also no mention made on orthopedic intervention. Tendon lengthening definitely also influences tone, albeit in a lesser degree, but they section the Golgi tendon organ that also affects afferent tone feedback neural fibres. This needs to be kept in mind and declared upfront. Botox may have effect up to 5 months, therefore you can not view the last dose at 8 weeks prior to evaluation, as no effect anymore. Also: Botox, even if in intervals do affect development in a patient - by reducing tone and therefore it may also reduce the longterm consequences of untreated spasticity in CP. In conclusion: the protocol will not adequately have 2 different groups to compare. SDR is viewed as aggressive therapy in the author's eyes, but they must remember that it reduces tone, in the same way as Botox, Baclofen and even tendon lengthening does. The major difference is the duration, cost and longterm sequelae from each procedure, but they can not be seen as aggressive, vs non-aggressive = physiological effect remains similar. I think that there needs to be revisions done to address the above concerns first. It will still be a very worthwhile exercise and I urge the authors to look at making the suggested changes and resubmitting as the results of a properly divided subject group comparison will be very useful for our understanding of CP management in terms of spasticity.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

This is an extremely interesting study protocol. The need for appropriately matched control groups in SDR outcome research and CP research more generally has been well highlighted in the recent literature and is well summarised in this manuscript. The study itself should add significantly to the current literature and the study protocol is certainly worthy of publication in its own right and would be of significant interest to researchers in this area as the difficulty in combining differing spasticity assessments in particular either across centres, or historically within a single centre, is common to many motion laboratories and centres. The authors have proposed a unique and novel method of dealing with combining different methods of assessing spasticity. As mentioned above many centres struggle in dealing with historical records of spasticity assessment which has changed over time and overall the proposed method looks very interesting and valid and will allow participants to be categorised as mild/moderate/severe.

- We thank the reviewer acknowledging the value of this work. Due to some confusion from Reviewer 2, and our concern that other readers would have the same confusion, we heavily reformatted the structure of how this protocol paper was presented. While our methods are the same, the way in which they are presented has changed. Instead of focusing on +/- SDR groups, we chose to present our study as a comparison of two ends of a heterogeneous spasticity management spectrum. On one end are individuals who underwent a highly-interventional management philosophy for spasticity. These individuals had spasticity reduced at an early age via SDR, and maintained with ongoing management throughout their childhood and adolescent years. We refer to them as “Yes-SDR”. On the other end are individuals who received a minimally-interventional strategy. We refer to them as “No-SDR”. We believe this reorganization of the same study will be well-received by our intended audience. Heavy edits were made, so please refer to the updated version of the protocol.

However, the questioned asked in the review form above is-

"Are the methods described sufficiently to allow the study to be repeated?" and the description given is not quite clear enough to say 'yes . However, I appreciate that this may well be much clearer when seeing it 'in action' in the actual study paper and is possibly part of the difficulty in writing/reviewing a protocol only.

- While we anticipate that the final manuscript will help with clarity, the point is well taken and efforts to bring clarity to the methods in this protocol manuscript were made throughout. Heavy edits were made through. Please refer to the updated version of the protocol.

The spasticity score for GIL and SPK is based on ashworth score and is clear. The paper then states that "subjects needed to have at least 3 of 4 spasticity measures on each side to be included". This isn't quite clear and assume this just means they needed an ashworth score recorded for at least 3/4 muscles on each side to be considered for inclusion in the study rather than a particular score signifying spasticity to be present before being included?

- You are correct; patients who had at least 3 of 4 muscles evaluated for spasticity at the time of a prior clinical visit were included. We made this choice to increase our sample size without compromising the integrity of our study design. It is possible that patients who only had 1 or 2 muscles tested per side, for reasons unknown to us, were not candidates for SDR. The methods have been updated for clarity and now read:

“At GIL and SPK, the spasticity score was defined as the mean of bilateral Ashworth scores for 4 muscles: hip adductors, hamstrings, vasti/rectus femoris, and ankle plantar flexors. Individuals needed to have at least 3 of 4 spasticity measures on each side to have their data included. We made

the assumption that missing only 1 spasticity measure per limb still allowed for an unbiased summary of overall spasticity, while maintaining an adequate sample size."

The score for SLC is less clear. Again, assume the statement that subjects need 3/5 measures on each side means they just needed 3/5 measures recorded rather than referring to a measure signifying spasticity present? Also, the method of scaling some of these measures on a zero to five range is not clear (particularly Duncan Ely, clonus, deep tendon reflex and ankle stretch). When referring to difference between ankle initial stretch and final range does that initial stretch mean a fast stretch? Just wondering also, why weight the Ashworth scores by factor of two? Is this a reflection of its relative importance compared to the others (which I'm not disagreeing with!) or for mathematical/data management reasons?

"At SLC, Ashworth scores were not always used to document spasticity. Therefore, the lower limb spasticity score was defined as a weighted average of up to five different bilateral spasticity measures. 1) mean modified Ashworth for hamstrings, vasti/rectus femoris, ankle plantar flexors, and ankle invertors, 2) Duncan Ely for rectus femoris, 3) beats of clonus, 4) difference between ankle initial stretch and final range of motion as a surrogate for ankle plantar flexor spasticity, and 5) deep tendon reflex. All scores were scaled to be on a standardized 0 to5 range, where 5 was the higher severity for the measure. Modified Ashworth scores were weighted by a factor of two due to their relative importance compared to other measures (i.e. more direct measure of spasticity and utilization by GIL and SPK)."

Overall, very interesting protocol of significant interest and benefit to CP researchers but bit more clarity needed on spasticity scoring proposed

- We thank the reviewer for his comments and are hopeful our substantial edits provide the needed clarity.

Reviewer: 2

Thank you for performing this valuable study. i think it may add greatly to our current understanding of the effect and place of SDR compared with other treatments.

My biggest concern: I am not sure from the protocol if you will have properly compared what you think you are comparing: SDR vs non-SDR on function, etc. The authors have to bare in mind that Botox, as well as ITB during the course of the child's life will affect tone, function and therefore, outcomes also. This is therefore not a SDR vs none study. In both groups tone are influenced by treatment. It is just the method that differs.

- We agree that our proposed study is not designed to test the effect of SDR versus no treatment study. It is not practical to be able to design such a study retrospectively, nor is it ethical to plan such as study prospectively. The goal of the present study is to assess a highly-interventional spasticity management strategy, one which includes SDR, versus a minimally-interventional strategy, which does not include SDR. While much of the literature we presented in the introduction was focused on outcomes of SDR, we tried to make our broader focus clear. To further improve the clarity, a large re-working of our protocol was performed. Please reference the new version.

- We agree that use of baclofen impacts tone and function, which is why patients who have used an ITB pump or oral baclofen for more than 1 year are excluded from participation. It is not

uncommon for patients to trial the use of oral or an ITB pump and then, for a variety of reasons, discontinue use. We see no reason that including individuals who have trialed baclofen for <1 year, whether they went on to receive an SDR or not, should be excluded. If fact, we believe that including these individuals captures the typical CP population we are trying to understand, therefore making our results more generalizable. Additionally, in order to not sacrifice the legitimacy of our study design, we believe that any residual effect on tone and function would be absent by the time of 6 months post-discontinuation. The half-life of Baclofen is hours, not weeks. Your comment on Botox will be further addressed below.

There is also no mention made on orthopedic intervention. Tendon lengthening definitely also influences tone, albeit in a lesser degree, but they section the Golgi tendon organ that also affects afferent tone feedback neural fibres. This needs to be kept in mind and declared upfront.

- Your point about tendon lengthening is relevant. While the direct impact of tendon lengthening has not been demonstrated, we agree with both the logic and the indirect evidence that it may have an impact. In our study, we are tracking intervening surgeries that each group receives. We will count and compare the number of tendon lengthening surgeries between the groups as part of the outcome. We believe that tendon lengthening (ortho surgery) is an “outcome” of poorly managing spasticity. At our center, we have observed (and published) that individuals who have undergone SDR require fewer soft tissue surgeries, such as tendon lengthening surgeries. Of course, this finding cannot adjusted for the idea that surgeons may orthopedically treat patients different based on their SDR history.

Botox may have effect up to 5 months, therefore you can not view the last dose at 8 weeks prior to evaluation, as no effect anymore. Also: Botox, even if in intervals do affect development in a patient - by reducing tone and therefore it may also reduce the longterm consequences of untreated spasticity in CP.

- According to the clinical trial data publically available on the impact of onabotulinumtoxinA on ankle plantarflexors, spasticity improvements peak by week 4, are maintained through week 8, then regress back to values clinically insignificant from baseline by week 12 post-dose. At our gait center, we opt to analyze patients as part of standard-of-care who are post-dose at least 8 weeks as our physicians don’t believe function is significantly impacted at this time. While our inclusion criteria includes the 8 week cutoff, it’s likely that the majority of our individuals will not have recently received injections. Average length from last injection will be summarized and reported in the final manuscript, allowing researchers and clinicians to make their own judgements on the legitimacy of our findings.

In conclusion: the protocol will not adequately have 2 different groups to compare.

SDR is viewed as aggressive therapy in the author's eyes, but they must remember that it reduces tone, in the same way as Botox, Baclofen and even tendon lengthening does. The major difference is the duration, cost and longterm sequelae from each procedure, but they can not be seen as aggressive, vs non-aggressive = physiological effect remains similar.

- While we disagree with the reviewer that SDR, Botox, Baclofen, and tendon lengthening reduces tone through the same mechanism, our intention with using the word “aggressive” had nothing to do with how well an SDR reduces tone. Instead, it had to do with its permanence. ITB and baclofen are reversible and, at least in our center, are often trialed by cautious families before opting for the irreversible surgical procedure of an SDR. The term aggressive also has to do with Gillette’s treatment philosophy toward spasticity management. We opt to eliminate it through SDR, ITB pumps, oral medications, and focal injections, with many of our patient receiving more than one treatment approach. However, we did a large re-working of our protocol. While the methods themselves are the same, we approached the introduction in a different way – now focusing on a “highly-interventional” vs. “minimally-interventional” approach. Please reference the new version of the protocol.

I think that there needs to be revisions done to address the above concerns first. It will still be a very worthwhile exercise and I urge the authors to look at making the suggested changes and resubmitting as the results of a properly divided subject group comparison will be very useful for our understanding of CP management in terms of spasticity.

VERSION 2 – REVIEW

REVIEWER	Rory O'Sullivan Central Remedial Clinic, Ireland
REVIEW RETURNED	28-Feb-2019

GENERAL COMMENTS	This re-submission has improved on the previous draft and re-framing the presentation as a long-term review of highly-interventional spasticity management versus a minimally-interventional strategy is very clear. As highlighted in previous review I think this study protocol is valuable in its own right as a method of combining spasticity measures and also the results of the proposed study will significantly add to the knowledge base and at least help answer the question-Does loss of spasticity matter? The added clarity in the description of the spasticity scores is noted. However, I am still unclear on the SLC score. Particularly, I am still unclear on how the proposed measures will be scaled to a standardized 0 to 5 range. For example are a certain number of beats of clonus allocated a score of 1 and a certain number score 5? or will the clonus scores for the population be divided into quintiles and scored 1-5 on that basis? Likewise with Duncan ely, reflex, ankle stretch... Am sure this is clear to the authors but it is not fully clear in the described methods and perhaps an added sentence or two is all that I needed to clarify. Very minor wording points- In the introduction; line 5-6. "sdr is the core of a highly-interventional management approach." Bit pedantic on my part I know but just wonder about 'core'. A centre could be highly interventional with botox for example without ever proceeding to SDR? Perhaps SDR is the endpoint or something like that? Introduction; lines 12-16. This section is not fully clear, particularly the sentence "many children and adolescents continue to receive ongoing anti-spasticity management". The location of this sentence and flow of the paragraph appears to be suggesting that on-going anti-spasticity management continues after SDR which I'm not sure is what the authors intend to suggest? To be honest, this sentence could just be removed altogether without changing the meaning and flow of that paragraph.

REVIEWER	Dr Johannes MN Enslin Red Cross War Memorial Hospital Department of Neurosurgery Cape Town South Africa
REVIEW RETURNED	06-Mar-2019

GENERAL COMMENTS	Thank you for doing the updates. I am looking forward to your findings. Please remember to carefully check previous management etc and report on these in the results.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

The added clarity in the description of the spasticity scores is noted. However, I am still unclear on the SLC score. Particularly, I am still unclear on how the proposed measures will be scaled to a standardized 0 to 5 range. For example are a certain number of beats of clonus allocated a score of 1 and a certain number score 5? or will the clonus scores for the population be divided into quintiles and scored 1-5 on that basis? Likewise with Duncan ely, reflex, ankle stretch...

Am sure this is clear to the authors but it is not fully clear in the described methods and perhaps an added sentence or two is all that I needed to clarify.

• Thank you for your additional feedback. We have created an Appendix to explicitly lay out the SLC methods.

Very minor wording points-

In the introduction; line 5-6. "sdr is the core of a highly-interventional management approach." Bit pedantic on my part I know but just wonder about 'core'. A centre could be highly interventional with botox for example without ever proceeding to SDR? Perhaps SDR is the endpoint or something like that?

• You are correct in that a botox-only treatment strategy could be considered highly-interventional by some. To clarify how we define ‘highly interventional’ in our study, the introduction has been reworded slightly:

An SDR is a surgical procedure that largely normalizes muscle spasticity and is the core of highly-interventional management approach described here.

Introduction; lines 12-16. This section is not fully clear, particularly the sentence "many children and adolescents continue to receive ongoing anti-spasticity management". The location of this sentence and flow of the paragraph appears to be suggesting that on-going anti-spasticity management continues after SDR which I'm not sure is what the authors intend to suggest? To be honest, this sentence could just be removed altogether without changing the meaning and flow of that paragraph.

• Actually, yes, we did intend to suggest that many individuals continue to receive anti-spasticity injections, even after receiving an SDR. According to our pilot work, individuals received an average of 8 anti-spasticity injections post-SDR (compared to 22 in our comparison group). We edited our introduction slightly to clarify our intention.

While SDR has been repeatedly shown to largely eliminate spasticity immediately and sustainably, questions remain regarding the other hypothesized benefits SDR aims to achieve. Additionally, despite undergoing an SDR, many children and adolescents continue to receive ongoing anti-spasticity management. As a result of the uncertainty surrounding spasticity management, some clinicians take an aggressive approach to reducing spasticity, while others opt for minimal spasticity reduction management.