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)	Efficacy of two injection-site localization techniques for botulinum toxin injections: a single-blind, crossover randomized trial protocol among adults with hemiplegia due to stroke.
AUTHORS	MOREL, Claire; HAURET, Isabelle; ANDANT, Nicolas; Pereira, Bruno; Bonnin, Armand; Coudeyre, Emmanuel

VERSION 1 - REVIEW

REVIEWER	Alessandro Picelli Neuromotor and Cognitive Rehabilitation Research Center; Department of Neurosciences, Biomedicine and Movement Sciences; University of Verona, Italy
REVIEW RETURNED	16-Mar-2016

GENERAL COMMENTS	ABSTRACT - Page 2, line 41: Is hemiplegia due to stroke? - Page2, line 44: please, use AbobotulinumtoxinA instead of Disport through the whole manuscript. - Re-write it according to the main corrections of the manuscript. KEY WORDS - Please, choose your keywords using the MeSH terms. - Please, do not use words included in the title. INTRODUCTION - General comment. I know that there is a scant literature about the comparison of BoNT-A injection techniques in adults. But this is your target in this protocol and you should focus on articles about adults. In my opinion to talk indifferently about adults and children is misleading and does not focus the reader on your aims. Please, try to modify the Introduction section according to my comment. - Page 5, lines 95-105: you should specify the type of patients you are talking about. For example, the 22% failure rate for the gastrocnemius reported by Ref. no. 2 regards children with PCI and not adults. Again, this is quite misleading considering that your protocol includes adult patients with stroke. You should report some data about injection precision in adult patients (for example, see the two articles published on the Journal of Rehabilitation Medicine by Picelli et al. in 2012 and 2014). METHODS - Please, specify the population you are talking about in the Objective section. - Randomization section: What about allocation? - Study description: you should give more information about the ES and US procedures (machine name, US frequency and probe, ES parameters, etc.)

	- Inclusion criteria: Why do you include only patients with a MAS score of 1+/4? - In my opinion a fixed dose of 1500U of AbobotulinumtoxinA is very high for the triceps surae with a MAS score of 1+/4. Why did you choose this dosage? - Patients: please report your inclusion/exclusion criteria in the text and delete Table 1. - Exclusion criteria: What about inclusion in other trials? - What about post-injection procedures? i.e. electrical stimulation, casting, taping, stretching, etc. - Page 14, line 292: the soles muscle is not bi-articular. - Blinding: How the success of blinding will be assessed? Educated guess? DISCUSSION - Page 18, lines 378-380: please give further information about fibrosis (i.e. Picelli et al. 2012; Pitcher et al. 2015). LANGUAGE: English needs minor revisions.

REVIEWER	Barbara Singer Professor Centre for Musculoskeletal Studies School of Surgery Faculty of Medicine, Dentistry and Health Science The University of Western Australia Australia
REVIEW RETURNED	21-Mar-2016

GENERAL COMMENTS	Firstly, the study received ethics approval in 2012 and 'data collection is ongoing' (P12). Consequently it does not meet BMJopen's criteria for a protocol paper. In addition, there are a number of concerns with the paper and these are outlined below: The following comments are made for the authors' consideration Abstract – it is unusual to refer to appendices in an abstract. Throughout the paper the authors may wish to consider 'people first language'. Particularly in the case of investigations involving community dwelling stroke survivors, it is inappropriate to talk about 'hemiplegic cases'. Rather the participants in this study are people who have hemiplegia following a stroke. The hypothesis may be that ultrasound is superior to other methods for guidance of intra-muscular botulinum toxin injection but in a number of places the authors use phrases like "prove the superior efficacy" (p2), "no study has proven the benefits…"(p4 and p9), which suggests a lack of scientific objectivity on the part of the authors. The goal of the study is to explore the outcomes associated with two methods for injection guidance in an unbiased manner. Introduction: There are a number of limitations on this study which are not reflected in the summary on p4. The most serious of these is the very likely inability to distinguish a difference between groups with the small number of subjects and a

	measure (Tardieu scale) with a large known inter-test variance. The study is underpowered for the relatively insensitive test and no data for reliability of the Tardieu for the person who is administering it are reported. The rationale for the sample size (based on 'recruitment capacity – p15) is not an appropriate justification. The main argument for the study appears to be that all of the three previous studies in this area have had methodological flaws and none have characterized their outcomes "in terms of efficacy and patient comfort" (P4). This brings up the key issue – what is 'efficacy' in the context of this study? The authors use the terms 'efficacy' and 'efficiency' interchangeably on P9. **Efficiency** is often used a surrogate term for **effectiveness** – defined as the extent to which an intervention can produce a benefit in 'real world' conditions, as opposed to **efficacy** – defined as the extent to which an intervention can produce a benefit in 'ideal', usually closely controlled conditions)(*eg* see Zwarenstein & Oxman (2006) J Clin Epidemiol; 59: 1125-6). <u>Methods</u> In Figure 1, the words 'ecography' and 'ultrasound' both used. There is a lot of overlap between sections in the methods which results in considerable repetition (eg info in last paragraph P12 and first paragraph P13 has already been provided). There are omissions in the way the cohort to be studied are described. What level of chronicity is to be included? (eg. Greater than 6 months post stroke is a commonly used definition of 'chronic' stroke). Do all participants have to be able to walk independently to be included? It is assumed that these participants meet institutional criteria for Botulinum toxin injection – presumably that they do not have fixed contracture, they have functional goals which can be addressed by improving their plantarflexor spasticity and they are expected to need more than one cycle of treatment (some further explanation of this latter point is essential, as this is key to the methodology). Clinically it would be difficult to justify providing a second injection without demonstrating that clearly identified functional goals have been met – as per best practice guidelines. At present no reference is made to this client centred aspect of the purpose of the intervention. It is unclear why it is necessary to set a dose limit for the Triceps surae muscles. The focus of the measurement is on gastrocnemius. Soleus is not a bi-articular muscle (P14). What if gastrocnemius is not a major contributor to gait dysfunction, will it still be injected? No rationale is provided for using the Modified Ashworth Scale as a part of the inclusion criteria, although the authors identify several limitations in this scale. The "more sensitive" (P14) Tardieu Scale is to be used as a primary outcome measure.

	An outdated version of the Tardieu Scale is cited. This scale was originally translated into English by Jean-Michel Gracies and the version used here was published in 2000. It's important to note that in subsequent versions of the Tardieu scale - the descriptor allocated to grade and angle are reversed. This is due to an error in the original paper. The correct version was published in a reliability study undertaken in a group of children with Cerebral Palsy. *Gracies JM et al (2010) Reliability of the Tardieu Scale for assessing spasticity in children with cerebral palsy. Arch Phys Med Rehabil;91(3):421-8.* This study also provides more recent data on expected inter-test variation of this measure. An additional reference which has been overlooked is: *Ben-Shabat E et al (2013) Intra- and interrater reliability of the Modified Tardieu Scale for the assessment of lower limb spasticity in adults with neurologic injuries. Arch Phys Med Rehabil;94(12):2494-501* The experience of the injector in using the two forms of guidance studied and the familiarity of the blinded assessor with the outcome measures are key to the success of the study but are not addressed in the protocol. The second paragraph on P12 is superfluous. It would suffice to say that the second injection procedure will be identical apart from the tracking method used. A justification for the timepoint of evaluation (1 month post injection) is required, especially since a number of studies have shown that functional gains may take longer to demonstrate. Will the participants receive any therapy following first or second injection? Increasingly evidence suggests that functional improvement is unlikely, especially in chronic stroke, in the absence of appropriately targeted therapy commenced soon after injection. The proposed statistical analyses are insufficiently detailed. Parametric statistics are inappropriate for analysis of most of the data that will be/ has been collected. Reference needs to be made to clinically significant change or at least minimally clinical important change for those measures for which these data are available (eg. Gait speed). Given the likely error inherent in the goniometric data from the Tardieu, and the emphasis of this study on 'efficacy' of botulinum toxin for plantarflexor spasticity, it is unlikely that the Tardieu will be a sufficiently sensitive tool to demonstrate the superiority of one form of injection guidance over another, and certainly not with the proposed sample size.

REVIEWER	Robert M West University of Leeds, UK
REVIEW RETURNED	24-May-2016

GENERAL COMMENTS	This is a well-written description of the trial and intended statistical

	analysis. I am a statistical reviewer. There is one important aspect however that is not covered in sufficient depth: how the analysis will proceed should there be drop out. The patients recruited to this study are stroke patients and stroke has a high mortality rate in the year post ictus. This trial seeks to retain stroke patients for 5 months. I would expect drop out from mortality and from patients withdrawing on health grounds. As a result the trial, although initially balanced, may well lose balance, and therefore efficiency of design. What are the implications for the power of the study? There is a description of analysis as a GLM with a random intercept for patient which is logical. How will the effect of drop out be assessed? It is unlikely that data will be missing at random. Is it possible that bias could be introduced? The sample size determination is well explained by the authors, especially the difficult circumstances with lack of information. For the primary outcome, what are the clinical implications for the effect sizes that can be predicted? (effect sizes for 0 and 0.5 correlation). I have waggled my foot and wondered what differences in angles might arise. I find it difficult to assess if the cited differences are optimistic. There are some minor errors: Page 10 Line 214 at might be inserted Page 13 Line 277 would an be better than the Page 17 Line 357 not in place of no Ref 16 has been cited a number of times. The Henzel study is detailed in the bibliography. On Page 17 Line 359 there is a citation of Feise [16]. Please check. Is the missing reference Feise RJ (2002) BMC Med Res Meth?

REVIEWER	Peter Watson Medical Research Council UK
REVIEW RETURNED	31-May-2016

GENERAL COMMENTS	Evaluating the effectiveness of two injection-site localization techniques for botulinum toxin injections, echography or electrostimulation: a single-blind, crossover randomized trial. bmjopen-2016-011751 Page 10. Given this is a crossover design and the group comparison of interest is electrostimulation and ultrasound I was confused to see echography mentioned under group 1 in the flow diagram as the treatment used for the first injection. Should not this be ultrasound to precipitate a paired t-test comparisons with electrostimulation assuming the absence of a carryover effect? Page 15. Could you please clarify that the power calculation on page 15 is for (as I understand it) sequence group (as defined in the flow diagram on page 10) and not the treatment group (defined as the electrical stimulation and ultrasound groups). The 8.5 I am assuming here is the pooled average treatment group standard deviation which is used in the denominator for evaluating Cohen's d which you

	have done. This results from as follows: 7.12/8.5 = d of 0.8, as quoted on line 52 of page 15, with type I error of 5% and power of 90% as quoted on page 15 (line 52) gives 15 patients in each sequence and 30 patients in total comparing the treatment groups in the crossover design. ie if we have two groups of form A and B and sequences A B and B A then we need 15 patients in the ordering groups A B and 15 with ordering B A giving (the same) 30 patients in each of the treatment groups A and B assuming an unpaired t-test (ie correlation=0 as mentioned on line 56 of page 18). Page 16, line 7 Similarly to be absolutely clear I think you could express the interaction of interest more clearly as a sequence by treatment group interaction which tests for the carryover effect. Page 16, lines 2-10. There is a second power calculation which seems to suggest that the difference between (paired) treatment groups decreases when there is a larger correlation between the treatment groups than when assuming the groups are independent. This cannot be true - the correlation influences the standard deviation of the difference between paired groups and not the difference itself. Please clarify this calculation or, preferably, remove it as the previous power calculation is sufficient. Page 16, line 23. What analysis will be used for the intention-to-treat in treating missing values. e.g. multiple imputation, last observation carried forward, mixed models? Is the intention-to-treat analysis performed for all comparisons e.g. non-crossover comparisons mentioned on line 14 of page 17 and when using a non-parametric test such as the Wilcoxon (line 54 of page 16)). One could, for example, use multiple imputation on a ranked t-test since ranked t-tests are equivalent to Mann-Whitney tests (Conover and Iman, 1981) and one can easily generalise a ranked t-test for use with multiple imputation since it has a estimate and standard error. Page 16, line 43. What is the difference between order processing and sequence? I thought there was only one sequence factor.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name
Alessandro Picelli
Institution and Country
Neuromotor and Cognitive Rehabilitation Research Center; Department of Neurosciences, Biomedicine and Movement Sciences; University of Verona, Italy.

Please state any competing interests or state ‘None declared’:
None declared.

Please leave your comments for the authors below
TITLE
- Please change the title including some information about the study population (adults with hemiplegia due to stroke?).
We thank the reviewer for this comment and we modified the title in the article. Indeed the study population includes adults with hemiplegia due to stroke.
Please find the new title: “Efficacy of two injection-site localization techniques for botulinum toxin

injections: a single-blind, crossover randomized trial protocol among adults with hemiplegia due to stroke."

ABSTRACT
- Page 2, line 41: Is hemiplegia due to stroke?
We apologize if this point was not sufficiently clear in the previous manuscript, we made the modifications page 2, line 40.

- Page2, line 44: please, use AbobotulinumtoxinA instead of Disport through the whole manuscript.
Modification had been made in the text, page 2 line 42.

- Re-write it according to the main corrections of the manuscript.

KEY WORDS
- Please, choose your keywords using the MeSH terms.
We thank the reviewer for this relevant comment and we modified the keywords. Please find the new keywords "echography, electrostimulation, soleus, gastrocnemius" page 1, line 21

- Please, do not use words included in the title.
Title and text had been modified in line with this request.

INTRODUCTION
- General comment. I know that there is a scant literature about the comparison of BoNT-A injection techniques in adults. But this is your target in this protocol and you should focus on articles about adults. In my opinion to talk indifferently about adults and children is misleading and does not focus the reader on your aims. Please, try to modify the Introduction section according to my comment.
- Page 5, lines 95-105: you should specify the type of patients you are talking about
The new version of the manuscript (line 98 page 5) makes the difference between the 2 populations.

For example, the 22% failure rate for the gastrocnemius reported by Ref. no. 2 regards children with PCI and not adults. Again, this is quite misleading considering that your protocol includes adult patients with stroke. You should report some data about injection precision in adult patients (for example, see the two articles published on the Journal of Rehabilitation Medicine by Picelli et al. in 2012 and 2014).
We read these articles with great interest and add references (line 102-104, page 5).

METHODS
- Please, specify the population you are talking about in the Objective section.
We add this precision. We clarify this point in the new manuscript (line 201-202, page 9).

- Randomization section: What about allocation?
Thanks a lot for this comment. We apologize for the lack of details. The random allocation process is performed using a computer software program from Stata that generates the random sequence (line 341 page 16).

- Study description: you should give more information about the ES and US procedures (machine name, US frequency and probe, ES parameters, etc.)
We add details about ES and US in the new manuscript (line 247-248, page 12) "We use Dantec Clavis® for electrostimulation injection and Sonosite Edge® with a 6-13 MHz probe for ultrasound injection."

- Inclusion criteria: Why do you include only patients with a MAS score of 1+/4?
We clarified this point in the manuscript (page 13, line 272). “The inclusion criteria are: age 18 to 80 years, hemiplegic sequelae of stroke, triceps surae spasticity evaluated >1 on the modified Ashworth scale and ability to provide written consent.”

- In my opinion a fixed dose of 1500U of AbobotulinumtoxinA is very high for the triceps surae with a MAS score of 1+/4. Why did you choose this dosage?
In fact it’s not a fixed dose but the maximum possible dosage. We change it in the manuscript (line 252, page 12).

- Patients: please report your inclusion/exclusion criteria in the text and delete Table 1.
We added inclusion criteria in the text and deleted Table 1 (line 271-281, page 13)

- Exclusion criteria: What about inclusion in other trials?
Inclusion in another study was an exclusion criteria, it’s precised line 279 p.13

- What about post-injection procedures? i.e. electrical stimulation, casting, taping, stretching, etc.
After each injection patients were told to continue not to change their usual care (physiotherapy, self-exercise program, lower limb orthosis) (line 264-265, page 13). “After each injection patients were told to continue their regular physiotherapy.”

- Page 14, line 292: the soles muscle is not bi-articular.
We agree that soleus is mono-articular, we modified this mistake in the text (page 14, line 296-297).”
This straight-knee assessment is relevant for simultaneously obtaining measurement of gastrocnemius muscles, which are bi-articular, and the soleus spasticity”.
- Blinding: How the success of blinding will be assessed? Educated guess?
Indeed an evaluator (physiotherapist) was blinded concerning the tracking technique (page 13 line 263-264) and patients were told not to tell him which kind of guiding injection technique they had.
“Each follow-up visit will be performed by an investigator (physiotherapist) blinded to the tracking technique. The patients are also told to hide the kind of injection they received.”

DISCUSSION
- Page 18, lines 378-380: please give further information about fibrosis (i.e. Picelli et al. 2012; Pitcher et al. 2015).
We thank the reviewer for this comment and add more information about fibrosis.

LANGUAGE: English needs minor revisions.
We thank the reviewer for this comment. The text has been translated by Cremer Consulting (http://www.cremerconsulting.com/fr/) and the final manuscript has been reviewed by Dr Bonnin who is fluent in english and native from Scotland.

Reviewer: 2
Reviewer Name
Barbara Singer
Institution and Country
Professor
Centre for Musculoskeletal Studies
School of Surgery
Faculty of Medicine, Dentistry and Health Science

The University of Western Australia
Australia

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below
Firstly, the study received ethics approval in 2012 and 'data collection is ongoing' (P12). Consequently it does not meet BMJopen's criteria for a protocol paper.
We thank the reviewer for his relevant comments. As we have seen in the instruction for author for study protocols in BMJ open web site: “Protocol manuscripts should report planned or ongoing research studies. If data collection is complete, we will not consider the manuscript.”

In addition, there are a number of concerns with the paper and these are outlined in the attached document.

Reviewer: 3
Reviewer Name
Robert M West
Institution and Country
University of Leeds, UK

Please state any competing interests or state ‘None declared’:
None

Please leave your comments for the authors below
This is a well-written description of the trial and intended statistical analysis. I am a statistical reviewer. There is one important aspect however that is not covered in sufficient depth: how the analysis will proceed should there be drop out.
We thank the reviewer for his relevant comments about analysis of drop out which will be an essential point for this study. We agree with this aspect and we thank for all remarks, which allowed us to improve our manuscript.

The patients recruited to this study are stroke patients and stroke has a high mortality rate in the year post ictus. This trial seeks to retain stroke patients for 5 months. I would expect drop out from mortality and from patients withdrawing on health grounds.
The patients are survivor and their stroke dates more than 6 months (page 10, line 210). We apologize if this point was not sufficiently clear in previous manuscript.

As a result the trial, although initially balanced, may well lose balance, and therefore efficiency of design. What are the implications for the power of the study?
We perfectly agree it but regardless the patients are survivor and their stroke dates more than 6 months, the power of study should be satisfactory.

There is a description of analysis as a GLM with a random intercept for patient which is logical. How will the effect of drop out be assessed? It is unlikely that data will be missing at random. Is it possible that bias could be introduced?
As discussed previously, drop-out bias should be minor. But according to reviewer’s comment and to proposition of reviewer #R4, we have added in revised manuscript this sentence: “A sensitivity analysis will be proposed to determine the nature of missing data and to apply the more appropriate

imputation approach like multiple imputation".

The sample size determination is well explained by the authors, especially the difficult circumstances with lack of information. For the primary outcome, what are the clinical implications for the effect sizes that can be predicted? (effect sizes for 0 and 0.5 correlation). I have waggled my foot and wondered what differences in angles might arise. I find it difficult to assess if the cited differences are optimistic.
We thank the reviewer for this relevant comment and apologize if this point was not sufficiently clear in the previous manuscript. Page 14 line 291-293: "The effect size calculation is based on an improvement of 7 angular degrees which is quite important considering regular ankle range of motion from 0 to 50°."

There are some minor errors:
Page 10 Line 214 at might be inserted. These aspects have been modified in revised manuscript
Page 13 Line 277 would an be better than the Page 17 Line 357 not in place of no These aspects have been reviewed in revised manuscript.
Page 17 Line 357 not in place of no Thanks for the comment. We have changed in the text.
Ref 16 has been cited a number of times. The Henzel study is detailed in the bibliography. On Page 17 Line 359 there is a citation of Feise [16]. Please check. Is the missing reference Feise RJ (2002) BMC Med Res Meth? Thanks for the comment. We have changed in the text.

Reviewer: 4
Reviewer Name
Peter Watson
Institution and Country
Medical Research Council
UK

Please state any competing interests or state 'None declared':
None declared

Please leave your comments for the authors below
Evaluating the effectiveness of two injection-site localization techniques for botulinum toxin injections, echography or electrostimulation: a single-blind, crossover randomized trial. bmjopen-2016-011751

Page 10. Given this is a crossover design and the group comparison of interest is electrostimulation and ultrasound I was confused to see echography mentioned under group 1 in the flow diagram as the treatment used for the first injection. Should not this be ultrasound to precipitate a paired t-test comparisons with electrostimulation assuming the absence of a carryover effect?
We thank the reviewer for his comment. We apologize for the vagueness in this methodology section, particularly in flow diagram. As it was detailed in manuscript, our main objective is to compare the efficacy of botulinum toxin injections in terms of guiding technique: ultrasound vs. electrical stimulation in patients with hemiplegia due to stroke. Also, the principal statistical analysis will concern the comparison between ultrasound vs. electrical stimulation. It would be more appropriate to define 'groups' like sequences as defined in statistical considerations (page 15, line 325): ultrasound then electrical stimulation and electrical stimulation then ultrasound. To compare the efficacy of botulinum toxin injections in terms of guiding technique, we have proposed to perform a repeated analysis of variance for cross-over designs taking into account the following effects: treatment group (ultrasound vs. electrical stimulation), sequence, carry-over and subject (as random-effect).

Page 15. Could you please clarify that the power calculation on page 15 is for (as I understand it) sequence group (as defined in the flow diagram on page 10) and not the treatment group (defined as the electrical stimulation and ultrasound groups). The 8.5 I am assuming here is the pooled average

treatment group standard deviation which is used in the denominator for evaluating Cohen's d which you have done. This results from as follows: 7.12/8.5 = d of 0.8, as quoted on line 52 of page 15, with type I error of 5% and power of 90% as quoted on page 15 (line 52) gives 15 patients in each sequence and 30 patients in total comparing the treatment groups in the crossover design. ie if we have two groups of form A and B and sequences A B and B A then we need 15 patients in the ordering groups A B and 15 with ordering B A giving (the same) 30 patients in each of the treatment groups A and B assuming an unpaired t-test (ie correlation=0 as mentioned on line 56 of page 18).
We thank the reviewer for his comment. We perfectly agree with the explanation and comment about sample size estimation. According to previous point, these aspects have been clarified in revised manuscript.

Page 16, line 7 Similarly to be absolutely clear I think you could express the interaction of interest more clearly as a sequence by treatment group interaction which tests for the carryover effect.
We thank the reviewer for his comment. According to previous points, these aspects have been clarified in revised manuscript.

Page 16, lines 2-10. There is a second power calculation which seems to suggest that the difference between (paired) treatment groups decreases when there is a larger correlation between the treatment groups than when assuming the groups are independent. This cannot be true - the correlation influences the standard deviation of the difference between paired groups and not the difference itself. Please clarify this calculation or, preferably, remove it as the previous power calculation is sufficient.
We perfectly agree that the second power calculation could be confusing. According to reviewer's comment, we have decided to remove it in revised manuscript.

Page 16, line 23. What analysis will be used for the intention-to-treat in treating missing values. e.g. multiple imputation, last observation carried forward, mixed models? Is the intention-to-treat analysis performed for all comparisons e.g. non-crossover comparisons mentioned on line 14 of page 17 and when using a non-parametric test such as the Wilcoxon (line 54 of page 16)). One could, for example, use multiple imputation on a ranked t-test since ranked t-tests are equivalent to Mann-Whitney tests (Conover and Iman, 1981) and one can easily generalise a ranked t-test for use with multiple imputation since it has a estimate and standard error.
We thank the reviewer for the very relevant comment. Regardless the patients are survivor and their stroke dates more than 6 months, drop-out bias should be minor. But according to reviewer's comment and to remark of reviewer #R3, we have added in revised manuscript this sentence: "A sensitivity analysis will be proposed to determine the nature of missing data and to apply the more appropriate imputation approach like multiple imputation".
By cons, it was not planned imputation for missing data for analyses mentioned in P17L14.

Page 16, line 43. What is the difference between order processing and sequence? I thought there was only one sequence factor.
Thanks for the comment. We have modified this point in revised manuscript, according to reviewer's comment. We have deleted the redundancy.

To date, only one comparative study focused on the protocol's topic has been published (17). Therefore, if scientific literature data provides information on the statistical variability of ankle dorsiflexion range of motion at slow and high speeds obtained using the Tardieu scale for patients having suffered from stroke (17), exhibiting a standard deviation around 8.5°, proposing an expected difference between the two treatments (ultrasound vs. electrical stimulation) proves challenging. In addition, in order to highlight the efficacy of botulinum toxin injections in terms of guiding technique,

namely ultrasound vs. electrical stimulation, sample size estimation was based on statistical power simulations in relation to recruitment capacity. To demonstrate a minimum difference of 7.12, with an effect size of 0.8, 15 patients per sequence (ultrasound stimulation then electrical vs. electrical stimulation then ultrasound) will be needed for a two-sided Type I error at 5%, a statistical power of 90%, and a correlation coefficient equal to 0, owing due to the cross-over design. For a more favorable correlation coefficient (for example 0.5), the difference expected with 30 subjects (15 patients per sequence) will be near to 5° (effect size of 0.6). Finally, if there would be an interaction effect “order processing x group”, only the results of the first period could be considered. Under the previous assumptions, notably 15 subjects per group, the expected difference between the two groups would be 10°. Statistical power estimations will be performed a posteriori on other components of the Tardieu scale: quality of muscle reaction (X) at slow and fast speeds, as well as angle of apparition of the muscle reaction (Y) at slow and fast speeds.
Statistical analysis will be performed on an intention-to-treat basis using the Stata software (Version 13, StataCorp, College Station, US) for a two-sided Type I error at α=5%. The patient characteristics will be described by numbers and associated percentages for categorical data. For quantitative parameters, mean (standard-deviation) or median (interquartile range) values will be calculated and presented according to statistical distribution. The assumption of normality will be studied by Shapiro-Wilk test. The primary endpoint, namely change in the ankle dorsiflexion range of motion at slow and high speeds obtained using the Tardieu scale, will be compared between the groups using a repeated analysis of variance (anova) for cross-over designs while taking into account the following effects: treatment group (ultrasound vs. electrical stimulation), order processing, sequence, subject (as random-effect), and carry-over. A particular focus will be given to the interaction “sequence x treatment (ultrasound vs. electrical stimulation)”. A sensitivity analysis will be proposed to determine the nature of missing data and to apply the more appropriate imputation approach like multiple imputation. If this test proves significant, the statistical analysis will only cover the first period of this cross-over study. The normality of residuals will be studied, as described previously. When endpoints do not assume the normality assumption, a non-parametric paired test like the Wilcoxon will be proposed. Analyses concerning the secondary endpoints (quality of muscle reaction [X] at slow and fast speeds, the angle of the muscle reaction apparition [Y] at slow and fast speeds, assessment of the triceps surae spasticity on the modified Ashworth scale, walking speed, and extent of pain at the injection site using a visual analogue scale, along with the duration of tracking and injection will be studied in a similar way as the primary endpoint. For categorical parameters, Stuart-Maxwell test for paired data or generalized linear mixed model taking into account the above-mentioned effects will be applied. Concerning non-crossover comparisons, usual statistical tests will be performed: Student t-test or Mann-Whitney test if the conditions of t-test are not met (normality or homoscedasticity verified using Fisher-Snedecor test) for quantitative parameters and Chi-squared test or Fischer’s exact test for categorical variables, if appropriate. As discussed by Feise (18), adjustment of Type I error (α) will not be proposed systematically, but on a case-by-case basis in the light of clinical considerations rather than statistical ones only.
Evaluating the effectiveness of two injection-site localization techniques for botulinum toxin injections, echography or electrostimulation: a single-blind, crossover randomized trial.

Reviewers comments:
The following comments are made for the authors’ consideration

Abstract – it is unusual to refer to appendices in an abstract.
We thank the reviewer for this comment and modified it in the abstract.

Throughout the paper the authors may wish to consider ‘people first language’. Particularly in the case of investigations involving community dwelling stroke survivors, it is inappropriate to talk about ‘hemiplegic cases’. Rather the participants in this study are people who have hemiplegia following a stroke.

We thank the reviewer for this comment and modified it in the text (for example page 1 line 89).

The hypothesis may be that ultrasound is superior to other methods for guidance of intramuscular botulinum toxin injection but in a number of places the authors use phrases like “prove the superior efficacy” (p2), “no study has proven the benefits…”(p4 and p9), which suggests a lack of scientific objectivity on the part of the authors. The goal of the study is to explore the outcomes associated with two methods for injection guidance in an unbiased manner. *
We thank the reviewer for this comment and we make modifications in the text page4 line 79 and page 9 line 194 “To our knowledge, no study has definitely yet successfully proven the benefits of ultrasound-guided botulinum toxin injections in terms of efficacy and patient comfort compared to other guiding techniques.”

Introduction:
There are a number of limitations on this study which are not reflected in the summary on p4.
The most serious of these is the very likely inability to distinguish a difference between groups with the small number of subjects and a measure (Tardieu scale) with a large known inter-test variance. The study is underpowered for the relatively insensitive test and no data for reliability of the Tardieu for the person who is administering it are reported. The rationale for the sample size (based on ‘recruitment capacity – p15) is not an appropriate justification.

The main argument for the study appears to be that all of the three previous studies in this area have had methodological flaws and none have characterized their outcomes “in terms of efficacy and patient comfort” (P4).

This brings up the key issue – what is ‘efficacy’ in the context of this study?
The authors use the terms ‘efficacy’ and ‘efficiency’ interchangeably on P9. Efficiency is often used a surrogate term for effectiveness – defined as the extent to which an intervention can produce a benefit in ‘real world’ conditions, as opposed to efficacy – defined as the extent to which an intervention can produce a benefit in ‘ideal’, usually closely controlled conditions) (eg see Zwarenstein & Oxman (2006) J Clin Epidemiol; 59: 1125-6).
Thanks for the comment. We have changed in the text and written ‘efficacy’ everywhere in the new manuscript for example line 200 page 9.

Methods
In Figure 1, the words ‘ecography’ and ‘ultrasound’ both used.
We have changed in the figure page 11 and use “ultrasound”

There is a lot of overlap between sections in the methods which results in considerable repetition (eg info in last paragraph P12 and first paragraph P13 has already been provided).
We thank the reviewer for the relevant comment. We modified it in the manuscript to make it as clear as possible. We apologize if this point was not sufficiently clear in previous manuscript (line 258, page 12).

There are omissions in the way the cohort to be studied are described.
What level of chronicity is to be included? (eg. Greater than 6 months post stroke is a commonly used definition of ‘chronic’ stroke).
The modification had been made modified in the study design line 210 p. 10. The patients are survivor and their stroke dates more than 6 months.

Do all participants have to be able to walk independently to be included?
We thank the reviewer for this comment: ambulation was assessed by using the Functional Ambulation Classification modified line 211-212, page 10. “Severity of the ambulation deficit was

considered by using the Functional Ambulation Classification modified and stratification was made on ambulation"

It is assumed that these participants meet institutional criteria for Botulinum toxin injection – presumably that they do not have fixed contracture, they have functional goals which can be addressed by improving their plantarflexor spasticity and they are expected to need more than one cycle of treatment (some further explanation of this latter point is essential, as this is key to the methodology). Clinically it would be difficult to justify providing a second injection without demonstrating that clearly identified functional goals have been met – as per best practice guidelines. At present no reference is made to this client centred aspect of the purpose of the intervention.
We thank the reviewer for this comment. In fact all the patients included in this study have already been injected Abobotulinum toxin with success and due to clinical assessment they need new Abobotulinum toxin.

It is unclear why it is necessary to set a dose limit for the Triceps surae muscles.
We have to set a dose limit (500 UI) for the triceps surae to have a good reproductibility.

The focus of the measurement is on gastrocnemius. Soleus is not a bi-articular muscle (P14)
We thank the reviewer for the relevant comment and we modify this mistake in the text (page 14, line 296).

What if gastrocnemius is not a major contributor to gait dysfunction, will it still be injected?
In our population gastrocnemius is a major contributor to gait dysfunction; the spastic equinus generate difficulties for balance, gait and also foot wear.

No rationale is provided for using the Modified Ashworth Scale as a part of the inclusion criteria, although the authors identify several limitations in this scale.
We thank the reviewer for this comment. The modified Ashworth Scale is the scale of reference in spasticity according to the AFFSAPS recommendations
"RBP/ Argumentaire/traitements médicamenteux de la Spasticité - a79f07eee915181bc9ae4e506140cecb.pdf [Internet]. [cité 30 juin 2016]. Disponible sur: http://ansm.sante.fr/var/ansm_site/storage/original/application/a79f07eee915181bc9ae4e506140cecb.pdf"
Furthermore this scale has good intra- and interrater reliability when used by trained medical professions Please find the article : "Brashear A, Zafonte R, Corcoran M, Galvez-Jimenez N, Gracies J-M, Gordon MF, et al. Inter- and intrarater reliability of the Ashworth Scale and the Disability Assessment Scale in patients with upper-limb poststroke spasticity. Arch Phys Med Rehabil. oct 2002;83(10):1349-54."
The "more sensitive" (P14) Tardieu Scale is to be used as a primary outcome measure.
An outdated version of the Tardieu Scale is cited. This scale was originally translated into English by Jean-Michel Gracies and the version used here was published in 2000. It's important to note that in subsequent versions of the Tardieu scale - the descriptor allocated to grade and angle are reversed. This is due to an error in the original paper. The correct version was published in a reliability study undertaken in a group of children with Cerebral Palsy.
Gracies JM et al (2010) Reliability of the Tardieu Scale for assessing spasticity in children with cerebral palsy. Arch Phys Med Rehabil;91(3):421-8.
We thank the reviewer for this very relevant comment and we modified it (cf Appendix 2, page 27)

This study also provides more recent data on expected inter-test variation of this measure.

An additional reference which has been overlooked is: Ben-Shabat E et al (2013) Intra- and interrater reliability of the Modified Tardieu Scale for the assessment of lower limb spasticity in adults

with neurologic injuries. Arch Phys Med Rehabil;94(12):2494-501
We add this reference (line 302-304, page 14).

The experience of the injector in using the two forms of guidance studied and the familiarity of the blinded assessor with the outcome measures are key to the success of the study but are not addressed in the protocol.
We specified in the manuscript that the injector and the blinded assessor have an experience > 3 years as it is written in the CONSORT Statement (line 249, page 12) (http://www.consort-statement.org/Search?q=non+pharmacologocal)

The second paragraph on P12 is superfluous. It would suffice to say that the second injection procedure will be identical apart from the tracking method used.
We thank the reviewer for the relevant comment and erase the second paragraph on page 12.

A justification for the timepoint of evaluation (1 month post injection) is required, especially since a number of studies have shown that functional gains may take longer to demonstrate.
We thank the reviewer for this comment. Most of the studies we mentioned in the manuscript evaluated the efficiency of the toxin at one month post-injection.

Picelli A, Lobba D, Midiri A, Prandi P, Melotti C, Baldessarelli S, et al. Botulinum toxin injection into the forearm muscles for wrist and fingers spastic overactivity in adults with chronic stroke: a randomized controlled trial comparing three injection techniques. Clin Rehabil. mars 2014;28(3):232-42.
Picelli A, Tamburin S, Bonetti P, Fontana C, Barausse M, Dambruoso F, et al. Botulinum toxin type A injection into the gastrocnemius muscle for spastic equinus in adults with stroke: a randomized controlled trial comparing manual needle placement, electrical stimulation and ultrasonography-guided injection techniques. Am J Phys Med Rehabil Assoc Acad Physiatr. nov 2012;91(11):957-64.
Will the participants receive any therapy following first or second injection?
We thank the reviewer for this comment. After each injection patients were told to continue their regular physiotherapy (line 264-265, page 13). “After each injection patients were told to continue their regular physiotherapy.”

Increasingly evidence suggests that functional improvement is unlikely, especially in chronic stroke, in the absence of appropriately targeted therapy commenced soon after injection.

The proposed statistical analyses are insufficiently detailed. Parametric statistics are inappropriate for analysis of most of the data that will be/ has been collected.
We thank the reviewer for the comment and we are sorry if the Statistical Section seemed insufficiently detailed. Statistical plan analysis has been proposed by the senior biostatistician associated to this project (Bruno Pereira, 210 articles referenced in PubMed). The statistical reviewers R3 and R4 have appreciated our proposition: “This is a well-written description of the trial and intended statistical analysis.” As it was indicated in Statistical Section, usual statistical tests will be applied according to assumptions of parametric tests and to statistical distribution of variables studied in this work. For example, it was clearly detailed: “When endpoints do not assume the normality assumption, a non-parametric paired test like the Wilcoxon will be proposed.”

Reference needs to be made to clinically significant change or at least minimally clinical important change for those measures for which these data are available (eg. Gait speed)
We thank the reviewer for this relevant comment. In fact we set a minimally clinical important change for Tardieu Scale at 7 angular degrees and an improvement of 0.2 meters/second of gait speed.
For the gait speed: Lewek, Reliability of spatiotemporal Asymetry During Overground Walking for

Individuals Following Chronic Stroke, JNPT, septembre 2011
For the Tardieu Scale the minimally clinical important change differs according to studies (4 to 19°) Haugh, a systematic review of the Tardieu Scale for measurement of spasticity, Disability and Rehabilitation, August 2006
Our study is based on the values of plantar flexion of a study of the Annals of Academy of Neurology: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3098519/table/T0002
Page 14 line 290-294: “The effect size calculation is based on an improvement of 7 angular degrees which is quite important considering regular ankle range of motion from 0 to 50°. Concerning the gait analysis, an improvement of 0.2 m/s of the gait speed is considered as a minimally clinical important change. “

Given the likely error inherent in the goniometric data from the Tardieu, and the emphasis of this study on ‘efficacy’ of botulinum toxin for plantarflexor spasticity, it is unlikely that the Tardieu will be a sufficiently sensitive tool to demonstrate the superiority of one form of injection guidance over another, and certainly not with the proposed sample size
We thank the reviewer for this comment. Tardieu scale is actually considered as a reference tool to assess spasticity. Please find bellow the explanation for the sample size page 15 lines 326-331: To demonstrate a minimum difference of 7.12, with an effect size of 0.8, 15 patients per sequence (ultrasound stimulation then electrical vs. electrical stimulation then ultrasound) will be needed for a two-sided Type I error at 5%, a statistical power of 90%, and a correlation coefficient equal to 0, owing due to the cross-over design. For a more favorable correlation coefficient (for example 0.5), the difference expected with 30 subjects (15 patients per sequence) will be near to 5° (effect size of 0.6).

Please find below the manuscript with the revisions:

Efficacy of two injection-site localization techniques for botulinum toxin injections: a single-blind, crossover randomized trial protocol among adults with hemiplegia due to stroke.

Claire Morel1,2, Isabelle Hauret3, Nicolas Andant4, Bruno Pereira4, Armand Bonnin1,2 and Emmanuel Coudeyre1,2,5

1 Service de Médecine Physique et de Réadaptation; CHU Clermont-Ferrand, 58 rue Montalembert, 63 000 Clermont-Ferrand, France
2 Université Clermont Auvergne, Clermont-Ferrand, France
3 Centre médical Etienne Clémentel, 63530 Enval, France
4 Biostatistics Unit, Délégation Recherche Clinique & Innovation (DRCI), CHU Clermont-Ferrand, 63 000 Clermont-Ferrand F-63003;
5 INRA, Unité de Nutrition Humaine (UNH, UMR 1019), CRNH Auvergne, Clermont-Ferrand, France
Corresponding author:
Emmanuel COUDEYRE, Service de Médecine Physique et Réadaptation, CHU Clermont-Ferrand, Hôpital Nord,
Route de Chateaugay, F-63118 Cébazat, France
ecoudeyre@chu-clermontferrand.fr

Key words: echography, electrostimulation, soleus, gastrocnemius

Word count: 5104 words

ABSTRACT

Introduction
Botulinum toxin injections are an effective treatment for limb spasticity following stroke. Different

tracking techniques are used for this purpose: palpation, electrostimulation, electromyography and ultrasound. Yet very few studies have compared these different techniques, and none has successfully proven the superior efficacy of ultrasound-guided injections compared to another tracking method. The primary objective of our study was therefore to compare the efficacy of botulinum toxin injections depending on the tracking technique used: ultrasound versus electrostimulation.

Methods and analysis
This is a clinical, single-center, prospective, interventional, single-blind, crossover randomized trial. In total, 30 patients aged between 18 and 80 years old presenting with triceps surae spasticity (evaluated >1 on the modified Ashworth scale) associated with hemiplegia sequelae due to stroke will be included. The patients will be selected among those who attend for consultation the Physical Medicine and Rehabilitation Department of the Clermont-Ferrand University Hospital. One group will receive the abobotulinumtoxinA (BoNT-A) injection guided by electrostimulation then ultrasound, the second group's botulinum toxin injections will be guided by ultrasound then electrostimulation. For each patient, the duration of study participation is 5 months. The primary endpoint is variation in passive ankle dorsiflexion range of motion at slow and high speeds (Tardieu scale) with the knee straight.

Ethics and dissemination
This study has ethical approval form the CPP of Rhônes-Alpes region. Results will be published in a peer-reviewed journal.

Trial registration
ClinicalTrials.gov Identifier, NCT01935544
Strengths and limitations of this study:
The management of muscle spasticity proves to be a major challenge in hemiplegia following a stroke, with botulinum toxin injections constituting the first-line treatment for local or loco-regional spasticity.
Concerning the limitations, it is a prospective study with inherent risks due to this type of studies such as lost to follow up bias.

1) INTRODUCTION

The management of muscle spasticity is a major challenge in hemiplegia following a stroke, with botulinum toxin injections constituting the first-line treatment for local or loco-regional spasticity (1).

Yet there is a range of techniques involving different methods for both injection and tracking. The most commonly-used tracking techniques are, anatomy palpation, electrostimulation, electromyography, and, more recently, ultrasound. Palpating children to guide injection is not reliable, particularly when the deeper muscles are concerned (e.g., only 12% successfully-positioned injections in the tibialis posterior and flexor carpi ulnaris), but also for more superficial muscles (22% failure rate for the gastrocnemius) (2). For children, injection guided via electromyogram (EMG) is not always appropriate, when there is difficulty obtaining active or passive muscle activation, to differentiate muscular activation of a specific muscle from surrounding muscles (3). In addition, with this technique, there is no correlation between the extent of spasticity and muscular activity (3). One article (4) shows that neither manual needle placement nor electrical stimulation is wholly accurate to inject gastrocnemius muscle of adults with spasticity.
Tracking via ultrasound is widely used in other indications, such as infiltrations in the locomotor system (particularly the tendons and joints) (5) or anesthetic nerve blocks (6). The primary advantages of ultrasound-guided botulinum toxin injection are that tracking is painless (3), fast (7), more precise (3), and thus safer, avoiding complications associated with subcutaneous, intravascular,

or too-deep injections (8).

A comparative study has assessed the efficacy of ultrasound-based tracking with an electrostimulation-based technique (9). The authors evaluated 32 children presenting with cerebral palsy sequelae, who were divided into two groups. All received botulinum toxin injection into the gastrocnemius, which was guided by either ultrasound or electrostimulation depending on the group. The techniques were evaluated based on three different scales: the Ashworth (Appendix 1), Tardieu (Appendix 2) (10), selective motor control (SMC), and Physician Rating scales. The authors observed a non-significant improvement in spasticity, assessed by the Ashworth and Tardieu scales, at 3 months post-injection, in the group treated with ultrasound-guided injections. In contrast, the electrostimulation-guided group showed non-significant improvement in motor control of the antagonistic muscles. The only significant differences revealed were improvements in walking pattern and foot-to-ground contact in the ultrasound-guided group. Nonetheless, the numerous controversial methodological choices made by the authors limited the relevance of these results.

There have also been two comparative studies evaluating the efficacy of ultrasound- guiding with that of techniques using electrostimulation or anatomy palpation (11) (12).

The first (11), conducted in 2012, compared these three injection-guiding techniques in the lower limbs. The trial involved 49 patients presenting with lower-limb spasticity following stroke, who were randomized into three groups, the first receiving injections guided by anatomy palpation, the second by electrostimulation, and the last by ultrasound. All received a botulinum toxin injection into the gastrocnemius, administered by the same physician. The investigator, who was blinded to the injection type, evaluated each patient on inclusion and at 1 month. The patients were forbidden from undergoing any form of physical therapy within the 3 months preceding the study and during its entirety. Ashworth and Tardieu scale results were assessed for all, along with passive dorsiflexion of the foot. The authors reported significantly improved passive dorsiflexion of the foot in the ultrasound-guided injection group compared to the electrostimulation-guided group. Moreover, the Ashworth scale results were significantly improved 1 month following botulinum toxin injection in the ultrasound-guided group compared to the group where anatomy palpation was used.

The second study (12) was conducted in 2013 and assessed upper-limb spasticity in 60 patients who had suffered from strokes. As in the above-described study, these patients were randomized into three groups of 20 each in order to compare the three injection-guiding techniques: ultrasound, electrostimulation, and anatomy palpation. Two injections were administered in at least two of the following muscles: flexor carpi ulnaris, flexor carpi radialis, flexor digitorum superficialis, and the flexor digitorum profundus. The same physician, experienced with using botulinum toxin under ultrasound-guiding, administered all the injections. An investigator who was blinded to the injection type assessed each patient at the beginning and 4 weeks into the study. The patients were forbidden from undergoing any type of physical therapy in the 3 months preceding the study and during its entirety. The Ashworth and Tardieu scale results were assessed, along with passive dorsiflexion of the wrist and fingers. One month following injection, the modified Ashworth scale scores significantly improved in the group having undergone ultrasound-guided injection compared to the group tracked using anatomy palpation, as did the Tardieu scale scores and passive mobilizations. In contrast, the authors found no significant differences between ultrasound- and electrostimulation-guiding for the different evaluations.

In both of these studies, the authors described limitations consisting of the absence of functional evaluation of the upper or lower limbs, owing to the short follow-up rendering this assessment difficult to implement, as well as of the injections being administered by only one physician experienced with ultrasound-guided injection. The authors also indicated that body mass index (BMI) was not taken into account in their studies, despite obesity potentially constituting a limitation to the accurate assessment of anatomical landmarks.

In a literature review (13), all four guiding techniques (anatomy palpation, electromyography [EMG], electrostimulation, and ultrasound) were compared, with advantages and disadvantages outlined for each. The authors retrieved and analyzed 15 articles, concluding that injection guided by anatomy

palpation required no equipment and only a small-sized needle. Yet deep or slighter muscles were more difficult to access. In addition, while EMG enabled the toxin to be injected closest to the motor end-plate, this technique could, however, not guarantee that the needle was actually in the target muscle. As for the electrostimulation-guided technique, its primary advantage appeared to be its precise localization capacity. Despite this, it can take a long time to perform and require more training than the EMG and anatomy palpation techniques. Finally, ultrasound was found to enable the real-time visualization of the needle's progression while avoiding certain structures like blood vessels or nerves, among other advantages. In addition, the needle used in this technique was finer and thus less painful. On the other hand, this technique was highly dependent on the operator's skill, potentially requiring the presence of an assistant for beginners.
All in all, guiding injections by anatomy palpation thus appears to be the least precise technique. The other guiding techniques appear to offer superiority, in terms of precision and thus efficacy, although further studies must be conducted in order to determine which technique achieves the best clinical results.
Another literature review (14) evaluated the impact of the different injection-guiding techniques on the efficacy of botulinum toxin when treating not only spasticity but also dystonia. This review covered 10 studies, seven of which were randomized. The authors reported a high level of evidence (Grade A) that instrument-based guiding, i.e., ultrasound, electrostimulation or electromyography, was more effective than manual guiding in the treatment of upper limb spasticity, spastic equinus following stroke in adults, and cerebral palsy in children. The review's conclusions were that no instrument-based guiding technique proved superior to another. At the present time, no recommendation can be made in terms of choosing the optimal guiding technique, although ultrasound nevertheless appears to be more effective than electrostimulation in spastic equinus treatment following stroke in adults (passive mobilization of the ankle) (11).

No study has as definitely yet successfully proven the benefits of ultrasound-guided botulinum toxin injections in terms of efficacy and patient comfort compared to other guiding techniques.

2) METHODS/DESIGN

Objective

Our main objective is to compare the efficacy of botulinum toxin injections in terms of guiding technique: ultrasound vs. electrical stimulation in patients with hemiplegia due to stroke.

The secondary objective is to demonstrate that ultrasound-guidance is a less painful localization technique.

Study design

This prospective, randomized, single-center, single-blind, crossover study will be conducted in chronic stroke patients (>6 months) presenting with triceps surae spasticity. Severity of the ambulation deficit was considered by using the Functional Ambulation Classification modified and stratification was made on ambulation. The patients will receive two injections; each administered using a different guiding technique. Randomization will determine which technique will be used in the first and second instances. The patients will be selected among those who attend for consultation at the Physical Medicine and Rehabilitation Department of the Clermont-Ferrand University Hospital. The botulinum toxin injections and assessments will take place in the same department. The study will last 5 months for each patient. This study does not present a major risk for the subjects. The main potential disadvantages to the treatment are injection pain or side-effects from the botulinum toxin (increased motor deficit or dysphagia).

The study design is presented in Figure 1.

Figure 1: Flow diagram showing the different stages of the protocol

Randomization

The patients will be randomly assigned to one of the above-described groups by means of a Latin square design in order to balance out the group numbers.

Study description

The patients pre-selected during consultation at the Physical Medicine and Rehabilitation Department of the Clermont-Ferrand University Hospital will be handed a letter containing information on the study protocol. They will then have 1 month to grant their consent, should they wish, and be included at their next consultation.

The following data will then be collected for each patient: age, gender, time since stroke, side affected by the cerebral lesion, current treatments and dosages (for managing spasticity and pain), date of first botulinum toxin injection, as well as severity of deficit (functional walking scale).

The initial assessment of the patients included in the study will be performed just prior to the first injection. This evaluation will be both clinical (assessment of the triceps surae spasticity based on the Tardieu and modified Ashworth scales) and instrument-based (walking speed using GAITRite, CIR Systems Inc. Sparta, New Jersey, the USA).

The first injection will be administered in the outpatient clinic by a therapist with injection experience of over 3 years, guided using ultrasound or electrostimulation, depending on the group. We use Dantec Clavis® for electrostimulation injection and Sonosite Edge® with a 6-13 MHz probe for ultrasound injection. The clinical investigator will randomize the patients then administer the injection according to the guiding method assigned. In total, 500 units of BoNT-A (Dysport®) will be injected into four separate areas of the triceps surae to have a good reproductibility. Further injections will be administered into other muscle groups, if necessary. The total dose for this injection will be minus 1,500 BoNT-A units. Any pain experienced during the injection will be assessed by means of vertical visual analogue scales, and the time required for tracking and administering the injection will be recorded.

The second injection will be administered 4 months after the first, also in the outpatient clinic. The procedure will be identical to the first injection apart from the tracking method used on this occasion which will differ from that used for the first.

The two follow-up visits will take place 1 month after each botulinum toxin injection. Each patient will be asked to attend the clinic for consultation so as to allow the efficacy of the injection to be assessed. This assessment will be both clinical (assessment of the spasticity of the triceps surae by means of the Tardieu and modified Ashworth scales) and instrument-based (walking speed using GAITRite®). Each follow-up visit will be performed by an investigator (physiotherapist) blinded to the tracking technique. The patients are also told to hide the kind of injection they received. After each injection patients were told to continue their regular physiotherapy.

This study is actually ongoing and the investigators are currently still collecting data. The contents of the manuscript have not been submitted or published elsewhere.

Patients

The inclusion criteria are: age 18 to 80 years, hemiplegic sequelae of stroke, triceps surae spasticity evaluated >1 on the modified Ashworth scale and ability to provide written consent. The exclusion criteria are: injection of botulinum toxin dating from over 3 months, previous ultrasound-guided injection of botulinum toxin, swallowing impairment, ongoing anti-vitamin K (AVK) anticoagulation treatment with international normalized ration (INR) > 3 during one week before randomization, ongoing aminoglycoside treatment, general anesthesia with planned curare injection during study

participation, implant with a pacemaker, history of ankle arthrodesis, other contra-indication for botulinum toxin injection: myasthenia gravis, pregnancy or breast feeding and patient included in other trials. The indication for botulinum toxin injection to the upper limb will not constitute a non-inclusion criterion for this study.

Evaluation

The primary endpoint is variation in passive ankle dorsiflexion range of motion at slow and high speeds (Tardieu scale) while keeping the knee straight.
The procedure consists of assessing the angle at which resistance manifests, as well as the intensity of this resistance to mobilization at slow and fast speeds (15). The ankle dorsal flexion angle will thus be measured by means of a goniometer during passive manipulation of the ankle with the knee being kept straight, before and after treatment. For the Tardieu Scale the minimally clinical important change differs according to studies (16) the effect size calculation is based on an improvement of 7 angular degrees which is quite important considering regular ankle range of motion from 0 to 50° (17). Concerning the gait analysis, an improvement of 0.2 meters/second of the gait speed is considered as a minimally clinical important change (18).
This straight-knee assessment is relevant for simultaneously obtaining measurement of gastrocnemius muscles, which are bi-articular, and the soleus spasticity.
The Tardieu scale is more sensitive than the commonly-used modified Ashworth scale. The latter only consists of five stages, which does not always allow for treatment efficacy to be evaluated. Furthermore, this scale does not take into account the velocity factor during spasticity (16). Nevertheless, validation studies pertaining to the Tardieu scale and involving the adult population are scarce in the scientific literature. Moreover the Tardieu scale is reliable for assessing spasticity in lower limb muscles of adults with chronic neurologic injuries(19).
Assessing the difference in the range of motion between slow and fast speeds is relevant because this takes into account not only the spastic component but also any potential tendon retraction.
The principal evaluation criterion will be measured on the day of injection and at the Month 1 assessment.

The secondary endpoints are:
- other components of the "Tardieu scale": quality of muscle reaction (X) at slow and fast speeds, as well as angle of apparition of the muscle reaction (Y) at slow and fast speeds;
- assessment of the triceps surae spasticity on the modified Ashworth scale;
- walking speed;
- extent of pain at the injection site using a visual analogue scale;
- duration of tracking and injection.

Statistical considerations :

To date, only one comparative study focused on the protocol's topic has been published (20). Therefore, if scientific literature data provides information on the statistical variability of ankle dorsiflexion range of motion at slow and high speeds obtained using the Tardieu scale for patients having suffered from stroke (20), exhibiting a standard deviation around 8.5°, proposing an expected difference between the two treatments (ultrasound vs. electrical stimulation) proves challenging. In addition, in order to highlight the efficacy of botulinum toxin injections in terms of guiding technique, namely ultrasound vs. electrical stimulation, sample size estimation was based on statistical power simulations in relation to recruitment capacity. To demonstrate a minimum difference of 7.12, with an effect size of 0.8, 15 patients per sequence (ultrasound stimulation then electrical vs. electrical stimulation then ultrasound) will be needed for a two-sided Type I error at 5%, a statistical power of 90%, and a correlation coefficient equal to 0, owing due to the cross-over design. For a more favorable correlation coefficient (for example 0.5), the difference expected with 30 subjects (15

patients per sequence) will be near to 5° (effect size of 0.6). Finally, if there would be an interaction effect “order processing x group”, only the results of the first period could be considered. Under the previous assumptions, notably 15 subjects per group, the expected difference between the two groups would be 10°. Statistical power estimations will be performed a posteriori on other components of the Tardieu scale: quality of muscle reaction (X) at slow and fast speeds, as well as angle of apparition of the muscle reaction (Y) at slow and fast speeds.
Statistical analysis will be performed on an intention-to-treat basis using the Stata software (Version 13, StataCorp, College Station, US) for a two-sided Type I error at α=5%. The patient characteristics will be described by numbers and associated percentages for categorical data. For quantitative parameters, mean (standard-deviation) or median (interquartile range) values will be calculated and presented according to statistical distribution. The assumption of normality will be studied by Shapiro-Wilk test. The primary endpoint, namely change in the ankle dorsiflexion range of motion at slow and high speeds obtained using the Tardieu scale, will be compared between the groups using a repeated analysis of variance (anova) for cross-over designs while taking into account the following effects: treatment group (ultrasound vs. electrical stimulation), order processing, sequence, subject (as random-effect), and carry-over. A particular focus will be given to the interaction “sequence x treatment (ultrasound vs. electrical stimulation)”. A sensitivity analysis will be proposed to determine the nature of missing data and to apply the more appropriate imputation approach as multiple imputation. If this test proves significant, the statistical analysis will only cover the first period of this cross-over study. The normality of residuals will be studied, as described previously. When endpoints do not assume the normality assumption, a non-parametric paired test like the Wilcoxon will be proposed. Analyses concerning the secondary endpoints (quality of muscle reaction [X] at slow and fast speeds, the angle of the muscle reaction apparition [Y] at slow and fast speeds, assessment of the triceps surae spasticity on the modified Ashworth scale, walking speed, and extent of pain at the injection site using a visual analogue scale, along with the duration of tracking and injection will be studied in a similar way as the primary endpoint. For categorical parameters, Stuart-Maxwell test for paired data or generalized linear mixed model taking into account the above-mentioned effects will be applied. Concerning non-crossover comparisons, usual statistical tests will be performed: Student t-test or Mann-Whitney test if the conditions of t-test are not met (normality or homoscedasticity verified using Fisher-Snedecor test) for quantitative parameters and Chi-squared test or Fischer’s exact test for categorical variables, if appropriate. As discussed by Feise (21), adjustment of Type I error (α) will not be proposed systematically, but on a case-by-case basis in the light of clinical considerations rather than statistical ones only.

3) DISCUSSION

The various comparative studies currently available (9), (11), (12) have demonstrated that instrument-guided procedures, such as electrostimulation and ultrasound, improve the efficacy of botulinum toxin injections compared to that obtained by means of simple anatomy palpation, in line with current recommendations for good clinical practices.
Ultrasound enables us to visualize in real-time the needle’s progress, resulting in a precise localization of the target muscle, while avoiding certain structures like blood vessels and nerves. In addition, this technique allows a passive manipulation of the limb part under study in order to distinguish the muscular body of the target muscle from that of other adjacent muscular structures (11).

Ultrasound-guided botulinum toxin injection can be subject to the same limitations inherent to ultrasound itself. The technique is highly dependent on the skills of the operator, who needs to be experienced, thus requiring further investment in terms of training and equipment. Additionally, the structural evolution of spastic muscles, as fatty infiltration and, in particular, fibrous involution, alter the ultrasound features of the muscle, rendering it at times difficult to distinguish from the different adjacent muscles (22), (23).

The substantial cost of ultrasound equipment no longer appears to represent an obstacle to using this guiding technique. It is now possible to directly employ different ultrasound waves with a digital tablet, thus considerably reducing the equipment costs.

With regard to the guiding speed of the different techniques, the literature currently provides contradictory views. The Berweck team (7) demonstrated that the mean time of muscle localization and injection was only 5 seconds for superficial muscles and 30 seconds for deeper ones when using ultrasound. On the other hand, the 2010 Henzel study (22) reported an average increase of 5 to 10 minutes in procedure time when adding ultrasound-guiding to usual guiding techniques. If ultrasound-guided injection was concretely proven to be faster than other methods, this could represent a particular advantage for children and poorly compliant adults displaying low tolerance for procedures involving prolonged immobilization (11).

In this study, we hypothesize that botulinum toxin injections guided by ultrasound are more efficacious than those using electrostimulation, with the triceps surae spasticity as primary evaluation criterion. In addition, we also seek to prove that ultrasound-guided botulinum toxin injections are less painful than those administered using electrostimulation, and that the time needed for localizing and injecting is shorter for the former.

The expected benefit for the patient is thus a more efficacious injection and consequently reduced spasticity of the triceps surae. The benefits of ultrasound-guided injection compared to that of electrostimulation-guided consist of reduced tracking and injection times, in addition to reduced pain on injection.

This study's objective is to improve the techniques pertaining to guiding and injection. When injecting botulinum toxin, it is, in fact, all the more crucial to be as precise as possible in order to ensure the best efficacy in the target muscles while avoiding any unwanted-effects that could arise in relation with toxin diffusion or intravascular injection. For this reason, it is highly-desirable to use the most reliable guiding method possible. Furthermore, toxin injection can be a painful procedure, particularly for certain patients suffering from hyperesthesia or cognitive disorders, meaning a guiding technique that enables the highest tolerance is all the more crucial.

VERSION 2 – REVIEW

REVIEWER	Alessandro Picelli Neuromotor and Cognitive Rehabilitation Research Center, Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Italy
REVIEW RETURNED	13-Jul-2016

GENERAL COMMENTS	The Authors addressed all my previous concerns in a satisfactory way.

REVIEWER	Peter Watson Medical Research Council UK
REVIEW RETURNED	14-Jul-2016

GENERAL COMMENTS	Efficacy of two-injection-site localization techniques for botulinum toxin injections: a single-blind crossover randomized trial protocol among adults with hemiplegia due to stroke. bmjopen-2016-011751.R1

	There is still a confused description of the power calculation for the test of the difference between ultrasound and electrical stimulation means on pages 14 and 15. My concerns are over the rationale for the size of minimum differences quoted on pages 14 and 15, which minimum difference is of interest (to use in the power calculation) and indeed I am not sure if they are powering using an unpaired t-test or paired t-test. They should also state whether or not they are assuming a carryover effect in the power calculation since this influences how many people they would need. My suggestion below is that the authors motivate the use of a single difference between the ultrasound and electrical stimulation and quote this for the most conservative case of a carryover effect comparing these only at period 1 using an unpaired t-test. Using only period 1 scores they would need 26 people on ultrasound and another 26 people on electrical stimulation for their quoted effect size of 0.8 (page 14) with 80% power. If there is no carryover effect they will still be adequately powered as the sd of the difference will be less due to a non-zero correlation between the ultrasound and electrical stimulation as these scores will both be available for each subject since sd (paired difference) = sqrt(sd1^2 + sd2^2 - 2rho sd1 sd2) where rho is the correlation assumed to be zero if only using period 1 scores. I, do appreciate, however that this would require more people than the 15 people per sequence (line 50 on page 14) since there would be only 15 people in each group at period 1 (which would be analysed only in the event of a carryover effect) rather than 26 as powered above. A less conservative approach, therefore, if you think a carryover effect is unlikely would have the stated 30 patients for each treatment for the 0.8 effect size stated on page 14 with 80% power because you could then use data from both periods and you have 15 people at each period on each of one of the treatments giving 30 people in each group > 26 required (as above) still assuming a worse case scenario of no correlation between patient scores on ultrasound and electrical stimulation. Page 14, line 56 & page 15, lines 1-5. I find the mention of another effect size corresponding to a correlation of 0.5 confusing. This undermines the power calculation described on lines 50-56 of page 14 which is powered assuming a higher effect size (of 0.8) and will now not have the power to detect the lower effect size (of 0.6) which would require more people (N=23) in each sequence making a total of 46 people required if one assumed no correlation (as used in the first power calculation) between the two responses (ultrasound and electrical stimulation) on a subject. Such a conservative assumption of zero correlation between responses on a subject assumes an unpaired t-test is used. One should quote only one effect size which is the minimum effect size of interest, in this case this would appear to be 0.6. There is also no reference given for the minimum differences quoted of 7.12 (line 50) or the 5 degrees on line 5 of page 15 so it is not clear on what basis these are chosen. Incidentally should these differences both be expressed in degrees as they are presumably measuring the same thing on the same scale? I also point out that if a carryover effect is found and only the scores used at period 1 (as correctly inserted on line 36 of page 15) you would also be underpowered since you could only use half the observations so it should be mentioned that the power calculation assumes there is no carryover effect. For example, if we have 15

	with sequence U(ltrasound) followed by E(lectrical Stimulation) and 15 with sequence E U and there is no carryover we have a total of 30 in each treatment (U and E) summing over the two sequences but, in the event of a carryover effect, we would only be using only period 1 and would have only 15 people on Ultrasound and 15 on Electrical so would have the power to be able to only detect smaller differences. The safest power calculation would give the number of people required to detect a given minimum difference at period 1 only (ie for an unpaired t-test). You could, for example, state that you require 26 people having ultrasound and 26 people having electrical stimulation at period 1 to detect an effect size of 0.8 (as quoted on line 50 on page 14) for a two tailed-type I error of 5% and 80% power provided you explain on what basis you have decided that the effect size such as that on page 14 quoted as 0.8 (based upon a difference of 7.12) is the minimum difference of interest. Page 14, line 56. I am afraid I can't see why a correlation coefficient of 0 is a result of having a crossover design as seems to be implied here. The crossover design (provided there is no carryover effect) would actually induce a correlation between the ultrasound and electrical stimulation scores since each person would have both (ie they would be two repeated measures). Is this, therefore, referring to a correlation between two other things? Page 15 Moving the sentence on lines 32-34 to after the sentence on lines 34-38 would be much clearer in that the test of carryover would then be described in the sentence immediately before the description of what is done if such a carryover effect is found. Page 14, line 56. 'owing due to the cross-over design' should read 'owing to the crossover design' I note cross-over is hyphenated here on line 56 of page 14 but not elsewhere including, for example, in the title of the paper on page 1. Page 16, line 7. 'Fischer' should read 'Fisher'.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 4
Reviewer Name
Peter Watson
Institution and Country
Medical Research Council
UK

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below
Efficacy of two-injection-site localization techniques for botulinum toxin injections: a single-blind crossover randomized trial protocol among adults with hemiplegia due to stroke. bmjopen-2016-011751.R1

There is still a confused description of the power calculation for the test of the difference between ultrasound and electrical stimulation means on pages 14 and 15. My concerns are over the rationale for the size of minimum differences quoted on pages 14 and 15, which minimum difference is of

interest (to use in the power calculation) and indeed I am not sure if they are powering using an unpaired t-test or paired t-test. They should also state whether or not they are assuming a carryover effect in the power calculation since this influences how many people they would need.

My suggestion below is that the authors motivate the use of a single difference between the ultrasound and electrical stimulation and quote this for the most conservative case of a carryover effect comparing these only at period 1 using an unpaired t-test. Using only period 1 scores they would need 26 people on ultrasound and another 26 people on electrical stimulation for their quoted effect size of 0.8 (page 14) with 80% power. If there is no carryover effect they will still be adequately powered as the sd of the difference will be less due to a non-zero correlation between the ultrasound and electrical stimulation as these scores will both be available for each subject since sd (paired difference) = sqrt(sd1^2 + sd2^2 - 2rho sd1 sd2) where rho is the correlation assumed to be zero if only using period 1 scores. I, do appreciate, however that this would require more people than the 15 people per sequence (line 50 on page 14) since there would be only 15 people in each group at period 1 (which would be analysed only in the event of a carryover effect) rather than 26 as powered above. A less conservative approach, therefore, if you think a carryover effect is unlikely would have the stated 30 patients for each treatment for the 0.8 effect size stated on page 14 with 80% power because you could then use data from both periods and you have 15 people at each period on each of one of the treatments giving 30 people in each group > 26 required (as above) still assuming a worse case scenario of no correlation between patient scores on ultrasound and electrical stimulation.

Page 14, line 56 & page 15, lines 1-5. I find the mention of another effect size corresponding to a correlation of 0.5 confusing. This undermines the power calculation described on lines 50-56 of page 14 which is powered assuming a higher effect size (of 0.8) and will now not have the power to detect the lower effect size (of 0.6) which would require more people (N=23) in each sequence making a total of 46 people required if one assumed no correlation (as used in the first power calculation) between the two responses (ultrasound and electrical stimulation) on a subject. Such a conservative assumption of zero correlation between responses on a subject assumes an unpaired t-test is used. One should quote only one effect size which is the minimum effect size of interest, in this case this would appear to be 0.6. There is also no reference given for the minimum differences quoted of 7.12 (line 50) or the 5 degrees on line 5 of page 15 so it is not clear on what basis these are chosen. Incidentally should these differences both be expressed in degrees as they are presumably measuring the same thing on the same scale?

I also point out that if a carryover effect is found and only the scores used at period 1 (as correctly inserted on line 36 of page 15) you would also be underpowered since you could only use half the observations so it should be mentioned that the power calculation assumes there is no carryover effect. For example, if we have 15 with sequence U(ltrasound) followed by E(lectrical Stimulation) and 15 with sequence E U and there is no carryover we have a total of 30 in each treatment (U and E) summing over the two sequences but, in the event of a carryover effect, we would only be using only period 1 and would have only 15 people on Ultrasound and 15 on Electrical so would have the power to be able to only detect smaller differences. The safest power calculation would give the number of people required to detect a given minimum difference at period 1 only (ie for an unpaired t-test). You could, for example, state that you require 26 people having ultrasound and 26 people having electrical stimulation at period 1 to detect an effect size of 0.8 (as quoted on line 50 on page 14) for a two tailed-type I error of 5% and 80% power provided you explain on what basis you have decided that the effect size such as that on page 14 quoted as 0.8 (based upon a difference of 7.12) is the minimum difference of interest.

Page 14, line 56. I am afraid I can't see why a correlation coefficient of 0 is a result of having a crossover design as seems to be implied here. The crossover design (provided there is no carryover effect) would actually induce a correlation between the ultrasound and electrical stimulation scores

since each person would have both (ie they would be two repeated measures). Is this, therefore, referring to a correlation between two other things?

We thank the reviewer for the very helpful, detailed and relevant comment about sample size estimation. We perfectly agree that the description of the power calculation could be considered as confused. We apologize. Some difficulties of this confusion were partly explained by methodological recommendations requested by French authorities. This part has been modified and simplified according to our suggestions.
In this sense, we agree that correlation coefficient equals 0 is not realistic for such crossover design study, or except in the unique interest of the first period. Furthermore, according to our experience, it seems realistic to assume no carryover effect. Finally, as detailed in statistical considerations, paired test as repeated analysis of variance were proposed to take into account appropriately the specifications of this design.
To date, only one comparative study focused on the protocol's topic has been published (20). Therefore, if scientific literature data provides information on the statistical variability of ankle dorsiflexion range of motion at slow and high speeds obtained using the Tardieu scale for patients having suffered from stroke (20), exhibiting a standard deviation around 8.5°, proposing an expected difference between the two treatments (ultrasound vs. electrical stimulation) proves challenging. In addition, in order to highlight the efficacy of botulinum toxin injections in terms of guiding technique, namely ultrasound vs. electrical stimulation, sample size estimation was based on statistical power simulations in relation to recruitment capacity. To demonstrate a minimum difference of 7.12 between ultrasound and electrical stimulations, with an effect size of 0.8 (so, an expected standard deviation of difference at 8.9°), 15 patients per sequence (ultrasound stimulation then electrical vs. electrical stimulation then ultrasound) will be needed for a two-sided Type I error at 5%, a statistical power of 80%, an intra-individual correlation coefficient equals 0.5 (owing to the cross-over design) and no carryover effect assumed.
Concerning the minimum difference quoted of the 5 degrees on line 5 of page 15 that was a request of our internal review committee. The committee initially asked us to present 2 different hypotheses on a more or less favorable point of view. According to literature data and based on our clinical experience we finally chose to conserve only one hypothesis with 7.12° in the final version of the paper.

Page 15 Moving the sentence on lines 32-34 to after the sentence on lines 34-38 would be much clearer in that the test of carryover would then be described in the sentence immediately before the description of what is done if such a carryover effect is found.
We thank the reviewer for this comment and modified it in the text.

Page 14, line 56. 'owing due to the cross-over design' should read 'owing to the crossover design' I note cross-over is hyphenated here on line 56 of page 14 but not elsewhere including, for example, in the title of the paper on page 1.

We thank the reviewer for this comment and modified it. The term crossover is included in the title of the paper: "Efficacy of two injection-site localization techniques for botulinum toxin injections: a single-blind, crossover randomized trial protocol among adults with hemiplegia due to stroke."

Page 16, line 7. 'Fischer' should read 'Fisher'.

Thank you for this comment, the modification has been made.