PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The ASPIRE trial: Study protocol for a double-blind randomized controlled trial of aspirin for overheating during exercise in multiple sclerosis
AUTHORS	Kever, Anne; Nelson, Katherine; Aguerre, Ines; Riley, Claire; Boehme, Amelia; Lee, Nancy; Strauss Farber, Rebecca; Levin, Seth; Stein, Joel; Leavitt, Victoria

VERSION 1 – REVIEW

REVIEWER	A/Prof Rhonda Brown Research School of Psychology, Australian National University, Australia
REVIEW RETURNED	06-Jul-2020

GENERAL COMMENTS	Bmj-open-2020-039691: Study protocol for a double-blind randomized controlled trial of aspirin for overheating during exercise in multiple sclerosis: The ASPIRE trial. Thank you for the opportunity to review this protocol of a double-blind randomized controlled trial of aspirin for overheating during exercise in people with multiple sclerosis (pwMS). Introduction The protocol has been clearly described; and in most cases, the rationale for the study has been adequately referenced. The RCT is based on sufficient evidence showing that there are multiple physical, psychological, and cognitive benefits of exercise for MS patients; although it increases core body temperature, which in turn, decreases exercise performance and causes fatigue. Nevertheless, it is clear exercise needs to be encouraged in pwMS as they are less physically active than people without MS and they may avoid exercise due to overheating and exhaustion, although exercise is safe in this population. A clear rationale was provided for the use of drug therapy that has utility relative to non-drug cooling methods (which may be cumbersome, expensive, etc.). It is appreciated that exercise can increase BDNF production and decrease circulating pro-inflammatory cytokines in MS patients. However, it was not clarified exactly: (a) how the molecules are pertinent to MS; and, (b) how core and peripheral (i.e. tympanic) body temperature are related to each other. Method The study protocol is based on a double-blind crossover pilot RCT of aspirin vs. placebo in 12 participants, with large effect size of 1.45 (Cohen's d) showing that aspirin results in an increase in exercise performance and reduces exercise-induced overheating in pwMS. In

	the current trial, acetaminophen constitutes a third 'arm' of the study; which has a different antipyretic mechanism to aspirin. The rationale to include acetaminophen should be moved to the introduction and a reference should be provided for the assertion that it has antipyretic effects. The study design is strong. Each participant is randomly allocated to their first treatment, and then in later sessions, they receive the other two 'treatments' via counterbalancing; using standard timing and duration of the sessions (2-hours), a >=1-week washout between the steps, and exercise scheduled for 1-hour after peak serum levels of the drug. The drug dosages appears to be appropriate (e.g. aspirin dose can reduce fatigue in non-exercising MS patients). Indirect recruitment is being used to recruit potential participants. Study inclusion and exclusion criteria appear to be appropriate, including restricting participation to patients with RRMS. Procedurally, the randomisation procedure is appropriate, clearly described, and there will be testing of the extent of blinding in the study. Ambient test conditions will be held constant across the participants and test days; and, ambient temperature on the morning of each test day will be recorded. Exercise tests and measures will be conducted by an exercise physiologist. Ethical approval has been granted for the study, and so far, 34 participants have been recruited and 16 have completed participation. The stopping criterion (<40 RPM) appears to be appropriate as it is based on volitional exhaustion criteria. Nevertheless, a number of issues require clarification: (i) What happens if a participant relapses during the trial - presumably they will be excluded? If so, the estimated sample size will need to be altered to cover this eventuality (based on the estimated annual relapse risk of participants). (ii) A sample size of 55-60 assumes an effect size of around .7, which is a large effect not a medium one. Provide a reference supporting the estimate that 60% of participants screened will meet eligibility criteria. (iii) It is appropriate to ask participants not to eat for 2-hours prior to testing as eating can increase body temperature – but a reference for this assertion is required. (iv) A potential procedural problem is that baseline questionnaires are typically filled out by participants before the randomization procedure, not afterwards – is there a reason this wasn't done? (v) Why were the additional lab measures (e.g. HR, BP, RER, etc.) included in the study? (vi) Each participant will have an individualised progressive ramping applied to their exercise regimen during the testing, based on their baseline self-reported exercise frequency and intensity. Is it sufficient to use self-report (rather than lab data) for this purpose, especially as subjective reports of physical activity do not correlate well with objective laboratory (or accelerometry) data? (vii) The study questionnaire, which includes the HADS and FSS and a physical activity measure, is appropriate. However, details of the scale's item response structure and psychometrics is lacking. (viii) Why is 'hours slept in previous night' included as a study variable? Participants should be asked to verify the time they went to bed, fell asleep, woke up, and got out of bed, as well as the estimated sleep duration; although it may not be possible to alter the study questions at this stage. Further, why ask about the presence of illness in a family member; and, how are the 10-point VAS estimates of pain, fatigue, and sadness, and ambient test day temperatures to be handled in the

	analyses? If the additional variables are to be used as covariates in the planned analyses, a revised sample size calculation should be provided for an Analysis of Covariance in which the covariates are to be controlled. Further, regarding the analyses, it is presumed the aspirin data will be combined (across the 3 treatment groups) and that the acetaminophen and control data will be treated in the same way? If so, it may be necessary to control for the order in which participants received the three treatments. Although their allocation to the first treatment condition is randomized, their assignment to the second and third 'treatments' is not random.

REVIEWER	Cinda Hugos VA Portland Health Care System Oregon Health & Science University United States of America
REVIEW RETURNED	13-Jul-2020

GENERAL COMMENTS	The editor should decide if specialist statistical review is needed. See below for no responses above. Thank you for the opportunity to review this interesting study on the use of aspirin to manage overheating during exercise in people with MS (PwMS). Page 2 Abstract Line 10 You only mention ASA. Line 18 Now APAP is mentioned. See below. Line 20 State how long after medication administration exercise begins. The introduction seems like it could be organized more logically. For instance: general exercise benefits and source of heat, history of exercise in PwMS – bad now good, elevated body temp without exercise in PwMS, then summarize all with your pilot of aspirin and exercise in PwMS. Page 4 Introduction Line 34 Uhthoff's describes vision changes from heat. It is a well-described elicitation but not the only well-described elicitation. Line 37 The sentence beginning "Importantly…" is out of place. Line 49 How much is the "subsequent increase in core temperature"? And in the next sentence, how much does the internal body temperature need to rise for physical fatigue to increase and exercise performance to worsen? Is this in healthy people or PwMS? Line 51 What is the "slight increase in body temperature" in PwMS? Line 53 This needs a better tie in to the statement "perhaps due to increased basal body temperature…" Page 5 Line 5 You do not have the Schwid NASA cooling study referenced. Line 17 Suddenly you are adding a third arm to this ASA vs placebo study. What is the background for APAP?

	Line 21 You assume ASA and APAP are equivalent. Based on what information? Line 36 You are back to “an oral antipyretic” and not 2 agents. Line 41 ASA does have some side effects and may not be recommended for everyone. This sentence makes it sound perfectly safe for all in any circumstances. Line 45 How safe is APAP? Line 50 Above, APAP was the third arm. Now placebo is the third arm. Page 6 Exclusion criteria – sounds like ASA and APAP may not be perfectly safe as alluded to above. Page 7 Sample size determination – did your pilot also have 3 arms? Are you sure 54 people is adequate with 3 arms? Line 33 Discontinued patients or those that deviate from protocols should be captured somehow. Line 45 Do they receive their exercise test results as additional compensation? Do they eventually learn perhaps why they may have done better at 1 (or 2) sessions than at the other(s)? Page 8 Line 7 Who asks participants about blinding effectiveness? How are they asked, in person, on paper? Line 22 Do they consent before each exercise test? In addition to consenting to participate in the study? Line 33 What is the dose of APAP? How was it selected? Line 35 How is overencapsulation of APAP handled? Page 9 Line 9 Is this a standard ramping protocol? Does it serve to make the data more comparable between subjects in terms of exercise duration? It seems with an outcome of time to exhaustion, this graded increase based on reported activity/fitness level makes your data less interpretable. 10W/min is 2x the increase of the first group and 15W/min is 3x the increase of the first group. Please clarify. Line 46 Do you think they will be anxious about the maximal exercise test, especially the first one and perhaps the second and third, if they did not like the exhaustion experience the first time? How will you account for this on the HADS scale? Page 10 Lines 33-40 How are you controlling for recruitment bias of people interested and experienced in exercise? Line 52 Fix grammar/typo. Page 11 Line 9 Are you asking participants to not take any ASA or APAP while enrolled in this study? Line 14 What is the one primary outcome of this study? Page 12 Line 5 Fix grammar. Line 20 Cooling methods are not necessarily administered serially.

	Line 28 Please provide evidence for the “favorable safety profile” of ASA. Lines 28-32 As this sentence now reads, it is an overarching statement and not accurate. What about Ampyra? Line 34 What is the basis of your statement about extant studies? Line 39 You could link back to Page 3 and the statement about application to Pw progressive MS.

VERSION 1 – AUTHOR RESPONSE

Reviewer: #1

The reviewer refers to the protocol as a clearly described RCT that is based on sufficient evidence. The reviewer’s concerns have been addressed in the revised version of the manuscript.

1) Introduction: It is appreciated that exercise can increase BDNF production and decrease circulating pro-inflammatory cytokines in MS patients. However, it was not clarified exactly: (a) how the molecules are pertinent to MS; and, (b) how core and peripheral (i.e. tympanic) body temperature are related to each other.

Response: Thank you very much for noting our omission. Clarification of this point has been provided in the revised manuscript. Here we provide a comprehensive response to the reviewer: (a) Pertinence of cytokines and BDNF to MS: During exacerbations of MS disease activity, elevated levels of circulating proinflammatory cytokines are found in cerebrospinal fluid and serum. BDNF is one of the neurotrophic factors shown to play a key role in counteracting the damaging inflammatory process via regulation, repair, and regeneration of neurons. (b) The reviewer makes an important point that peripheral and core body temperature are different parameters. Our decision to measure only peripheral body temperature (which can be conveniently and non-invasively collected, to reduce patient burden) was based on our pilot trial, in which this measurement was shown to be responsive to aspirin pre-treatment (Leavitt et al 2018). We now note in the manuscript that tympanic measurement of peripheral body temperature represents a standard non-invasive method shown to best approximate core temperature (Chamberlain et al., 1995; Henker & Coyne, 1995).

2) Method:

(a) The rationale to include acetaminophen should be moved to the introduction and a reference should be provided for the assertion that it has antipyretic effects.

Response: The rationale has been moved to the introduction and a reference has been added:

Hersh, E. V., Moore, P. A. & Ross, G. L. Over-the-counter analgesics and antipyretics: A critical assessment. *Clin. Ther.* **22**, 500–548 (2000).

(b)The study design is strong […]. Nevertheless, a number of issues require clarification:

(i)What happens if a participant relapses during the trial - presumably they will be excluded? If so, the estimated sample size will need to be altered to cover this eventuality (based on the estimated annual relapse risk of participants)

Response: These are good points that have now been clarified in the manuscript, i.e., if a participant relapses, s/he will indeed be excluded. As mentioned, we expect 10% attrition, which accounts for both participant drop-out as well as exclusion due to relapses.

(ii) Provide a reference supporting the estimate that 60% of participants screened will meet eligibility criteria.

Response: This estimate was based on our pilot trial, a point which is now noted in the revised manuscript.

(iii) It is appropriate to ask participants not to eat for 2-hours prior to testing as eating can increase body temperature – but a reference for this assertion is required.

Response: The requested reference has been added.

(iv) A potential procedural problem is that baseline questionnaires are typically filled out by participants before the randomization procedure, not afterwards – is there a reason this wasn't done?

Response: Questionnaires were deployed after randomization for practical reasons. After treatment administration, participants wait 1-hour to allow for peak serum level to be reached. In order to use this time efficiently and to avoid prolonging the experimental session for another hour (thereby adding undue time burden to the participant), we administer baseline questionnaires during this waiting time. At the editor's discretion, we would be happy to add these details to the Methods section of the paper.

(v) Why were the additional lab measures (e.g. HR, BP, RER, etc.) included in the study?

Response: Additional physiological indices are collected to allow for exploration of ancillary factors on an exploratory basis. Even more importantly, these lab measures are also required for any maximal exercise

test to monitor and assure the safety of the participant as per American College of Sports Medicine (ACSM) Guidelines (Riebe et al., 2015). This is now noted in the manuscript.

(vi) Each participant will have an individualized progressive ramping applied to their exercise regimen during the testing, based on their baseline self-reported exercise frequency and intensity. Is it sufficient to use self-report (rather than lab data) for this purpose, especially as subjective reports of physical activity do not correlate well with objective laboratory (or accelerometry) data?

Response: According to the ACSM's Guidelines, ramping tests according to participants' self-reported current exercise intensities and frequency along with review of their medical history by a certified exercise physiologist is sufficient to inform whether it is safe to conduct a maximal exercise test and what exercise testing protocol/ramp increments to utilize.

The ramp protocol used our study is in complete accordance with the ACSM's Guidelines for Exercise Testing and Prescription, wherein it states: *Ideally, increments should be chosen so that the total test time ranges between 8 and 12 min, assuming the endpoint is volitional fatigue. For example, increments of 10-15 W/min can be used on the cycle ergometer for older individuals, deconditioned individuals, and patients with CVD or pulmonary disease* (Pescatello et al., 2015).

(vii) The study questionnaire, which includes the HADS and FSS and a physical activity measure, is appropriate. However, details of the scale's item response structure and psychometrics is lacking.

Response: We now provide references to the original literature supporting our selected scales.

(viii) Why is 'hours slept in previous night' included as a study variable? Participants should be asked to verify the time they went to bed, fell asleep, woke up, and got out of bed, as well as the estimated sleep duration; although it may not be possible to alter the study questions at this stage. Further, why ask about the presence of illness in a family member; and, how are the 10-point VAS estimates of pain, fatigue, and sadness, and ambient test day temperatures to be handled in the analyses? If the additional variables are to be used as covariates in the planned analyses, a revised sample size calculation should be provided for an Analysis of Covariance in which the covariates are to be controlled.

Response: These are important points, and we thank the reviewer for bringing them to our attention. If the focus of the study was sleep, pain, fatigue, or sadness we agree that we would be remiss by not comprehensively interrogating these behaviors. However, these variables are only collected in the context of the ASPIRE study to support exploratory analyses of variables very much ancillary to the main focus of the study. We now explicate this point in the revised manuscript.

(ix) Further, regarding the analyses, it is presumed the aspirin data will be combined (across the 3 treatment groups) and that the acetaminophen and control data will be treated in the same way? If so, it

may be necessary to control for the order in which participants received the three treatments. Although their allocation to the first treatment condition is randomized, their assignment to the second and third 'treatments' is not random.

Response: We agree with the reviewer that it is necessary to control for order effects. The allocation schedule was designed in such a way as to randomly assign participants across all 3 study visits (not just the first visit), as noted in the Methods section entitled: *Allocation, randomization, blinding, and data protection*. As noted in the Data Analysis section, we control for order effects in all analyses.

Reviewer # 2

1) Page 2 Abstract:
Line 10 You only mention ASA. Line 18 Now APAP is mentioned. See below

Response: Thank you for noting this oversight; APAP is now mentioned as our comparison drug at the beginning of the abstract.

Line 20 State how long after medication administration exercise begins.

Response: We have added the requested detail to the abstract.

2) The introduction seems like it could be organized more logically. For instance: general exercise benefits and source of heat, history of exercise in PwMS – bad now good, elevated body temp without exercise in PwMS, then summarize all with your pilot of aspirin and exercise in PwMS.

Response: We appreciate the reviewer's suggestions and have revised the introduction in a manner that we believe provides a clear context, sufficient scientific justification, and germane background for our study.

2) Page 4 Introduction
Line 34 Uhthoff's describes vision changes from heat. It is a well-described elicitation but not the only well-described elicitation.

Response: We agree with the reviewer. We have edited the introduction accordingly.

Line 37 The sentence beginning "Importantly…" is out of place.

Response: The introduction has been revised as noted above.

Line 49 How much is the “subsequent increase in core temperature”? And in the next sentence, how much does the internal body temperature need to rise for physical fatigue to increase and exercise performance to worsen? Is this in healthy people or PwMS?

Response: In the revised introduction, we have clarified these details while maintaining a focus on persons with MS.

Line 51 What is the “slight increase in body temperature” in PwMS?:

Response: References to the relevant literature have been provided. At the editor’s discretion we are happy to expound upon the reviewer’s requested details here or in the Discussion, although we have chosen to leave them out as we believe they may detract from the focus of this study.

Line 53 This needs a better tie into the statement “perhaps due to increased basal body temperature…”

Response: We have omitted this speculative statement.

Page 5
Line 5 You do not have the Schwid NASA cooling study referenced.

Response: Thank you for noting this. We are huge fans of Steven Schwid and his excellent work, and in fact, are especially sorry that due to his untimely passing we are unable to collaborate with him on his very promising body of work, which indeed informed the conceptualization of our study. The noted citation is now included in the introduction.

Line 17 Suddenly you are adding a third arm to this ASA vs placebo study. What is the background for APAP?

Response: Thank you for noting our omission; please see response to Reviewer 1 above.

Line 21 You assume ASA and APAP are equivalent. Based on what information?

Response: Please see details now provided in Introduction, as well as explication provided to Reviewer 1.

Line 36 You are back to “an oral antipyretic” and not 2 agents.

Response: We have added the following to clarify: … an oral antipyretic (i.e., ASA and APAP)

Line 41 ASA does have some side effects and may not be recommended for everyone. This sentence makes it sound perfectly safe for all in any circumstances.

Response: Thank you; we have revised our language to be more appropriately circumspect.

Line 45 How safe is APAP?

Response: We are happy to provide details of the respective safety profiles for ASA and APAP, at the editor’s discretion. Relevant citations are provided for the interested reader.

Line 50 Above, APAP was the third arm. Now placebo is the third arm.

Response: Thank you for noting this. We have revised for consistency.

Page 6
Exclusion criteria – sounds like ASA and APAP may not be perfectly safe as alluded to above.

Response: Thank you for pointing out this discrepancy, we have tempered our language as noted above and have added an additional reference to provide access to supplementary information.

Page 7
Sample size determination – did your pilot also have 3 arms? Are you sure 54 people is adequate with 3 arms?

Response: Our pilot trial did not have 3 arms, as noted in the Introduction. Our sample size determination for this trial was made taking into consideration three arms. This is now noted in the Sample size determination section of the Methods.

Line 33 Discontinued patients or those that deviate from protocols should be captured somehow.

Response: Thank you for noting this very important methodological consideration. We now include mention of this in the revised manuscript.

Line 45 Do they receive their exercise test results as additional compensation? Do they eventually learn perhaps why they may have done better at 1 (or 2) sessions than at the other(s)?

Response: As mentioned in the section *Ethics and Dissemination,* we convey (anonymized) group results of all of our research studies to patients from our center via a bi-annual newsletter that our MS Center publishes. Upon individual patient request, we would gladly share results of an individual patient's exercise test performance after the study is completed, i.e., all study participants have concluded all 3 study visits. Interestingly (and anecdotally), no participant from either our pilot trial nor the ongoing ASPIRE trial has requested this specific information. Although we did not consider including this level of detail in the manuscript, at the editor's discretion, we are happy to include these details.

Page 8
Line 7 Who asks participants about blinding effectiveness? How are they asked, in person, on paper?

Response: This information is collected via a paper questionnaire administered by the study coordinator at the end of each study visit. The specification "via paper questionnaire'' has been added to the manuscript.

Line 22 Do they consent before each exercise test? In addition to consenting to participate in the study?

Response: Consent is given once, at the beginning of the study (Visit 1).

Line 33 What is the dose of APAP? How was it selected?

Response: The dose of APAP is 650 mg (standard adult dose) for consistency with ASA. This is noted in the *Pharmaceutical pretreatment* section of the manuscript.

Line 35 How is overencapsulation of APAP handled?

Response: As noted in the *Pharmaceutical pretreatment* section of the manuscript,

overencapsulation of APAP is handled by the Columbia University Irving Medical Center research pharmacy, consistent with the handling of the ASA and placebo.

Page 9
Line 9 Is this a standard ramping protocol? Does it serve to make the data more comparable between subjects in terms of exercise duration? It seems with an outcome of time to exhaustion, this graded increase based on reported activity/fitness level makes your data less interpretable.

Response: As noted above, we employed a standard ramping protocol based on the American College of Sports Medicine (ACSM) Guidelines (Riebe et al., 2015), which serve as the current gold standard in exercise physiology. Given the within-subject design of our trial, we do not agree with the reviewer that our data will be less interpretable.

Line 46 Do you think they will be anxious about the maximal exercise test, especially the first one and perhaps the second and third, if they did not like the exhaustion experience the first time? How will you account for this on the HADS scale?

Response: This is an interesting idea that can be explored in our data as an exploratory analysis.

Page 10
Lines 33-40 How are you controlling for recruitment bias of people interested and experienced in exercise?

Response: We are recruiting based on the inclusion/exclusion criteria explicated in the Methods section of the manuscript. The reviewer will note that there is no item regarding interest or experience in exercise amongst these criteria, as we do not believe this to be pertinent to the overarching hypothesis of the study. We are happy to add a discussion of this point, upon request.

Line 52 Fix grammar/typo.

Response: Thank you, we have corrected the typo.

Page 11
Line 9 Are you asking participants to not take any ASA or APAP while enrolled in this study?

Response: Participants are asked to refrain from use of ASA or APAP within 24-hours of their study visit. This information has been included in the Methods section (Page10).

Line 14 What is the one primary outcome of this study?

Response: We now clarify our primary outcome in sentence 2 of the Data Analysis section.

Page 12
Line 5 Fix grammar.

Response: Thank you. We have revised the grammar.

Line 20 Cooling methods are not necessarily administered serially.

Response: We are unsure what the reviewer means by this. We would be happy to clarify or revise our language if some discrepancy exists in our description.

Line 28 Please provide evidence for the “favorable safety profile” of ASA.

Response: We have included a citation to bolster the safety profile of ASA.

Lines 28-32 As this sentence now reads, it is an overarching statement and not accurate. What about Ampyra?

Response: Thank you for this important point. We have revised our language to be more accurate.

Line 34 What is the basis of your statement about extant studies?

Response: The statement is speculative, which we have clarified in the revised version. We believe it introduces an important point that should be made.

Line 39 You could link back to Page 3 and the statement about application to Pw progressive MS.

Response: Thank you, this is a nice suggestion that now constitutes the final sentence of the paper.

VERSION 2 – REVIEW

REVIEWER	A/Prof Rhonda Brown Research School of Psychology Australian National University Canberra, Australia
REVIEW RETURNED	21-Sep-2020

GENERAL COMMENTS	Thank you for making the suggested changes to the manuscript. Prior to publication a single sentence should be added to the

	Method detailing that some of the data collected in the study will not be reported in this paper.

REVIEWER	Cinda Hugos VA Portland Health Care System Oregon Health & Science University United States
REVIEW RETURNED	13-Aug-2020

GENERAL COMMENTS	Thank you for addressing all my comments/concerns.