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)	Does cannabidiol reduce severe behavioural problems in children with intellectual disability? Study protocol for a pilot single-site phase I/II randomised placebo controlled trial
AUTHORS	Efron, Daryl; Taylor, Kaitlyn; Payne, Jonathan M; Freeman, Jeremy; Cranswick, Noel; Mulraney, Melissa; Prakash, Chidambaram; Lee, Katherine; Williams, Katrina

VERSION 1 – REVIEW

REVIEWER	Javier Fernández-Ruiz Faculty of Medicine, Complutense University, Madrid, Spain
REVIEW RETURNED	07-Nov-2019

GENERAL COMMENTS	This submission is the description of a proposal for a pilot clinical trial with the phytocannabinoid cannabidiol addressed to attenuate severe behavioral problems frequently occurring in children and adolescents with intelectual disability. The idea is interesting and the design of the clinical trial is, in general, correct. It is supported by the recent benefits found against seizures, but also against long-term behavioral deterioration, with this phytocannabinoid used in medicinal cannabis extracts or formulated as Epidiolex for infantile refractory epileptic syndromes. Anyway, a couple of aspects (the time of active treatment that could be longer and to add some missing inclusion criteria) are susceptible of being improved (see below). In addition, I also detected some problems in the manuscript that would require attention before being published. Specific comments: 1. As mentioned above, the major problem with the design of this clinical trial is the time for the active phase of treatment which may be too short to confirm a therapeutic effect, as it has been documented in previous trials. I would recommend a longer duration of the active phase. In addition, given the low number of patients, 5 to be treated with cannabidiol + 5 with placebo, why not using a crossed design with 50% of patients having active treatment + washout period + placebo, and the other 50% placebo + washout period + active treatment? This could be more informative with patients being the own controls during the placebo phase for the active treatment. 2. Other question in relation with the design is to include “no previous use of cannabis” in the inclusion criteria, as it is possible that families may have treated these children with uncontrolled preparations before being recruited for the trial. This should be also confirmed with a blood analysis.

	3. A comment on Epidiolex and its recent approval for infantile refractory epileptic encephalopathies is necessary, e.g. page 6, line 12. 4. The second paragraph in page 6 would need to be rewritten. It combines the action of cannabidiol with the endocannabinoid signaling, but cannabidiol is one of the most atypical cannabinoids in relation with its activity at the classic elements of the endocannabinoid signaling. Most of its targeted proteins and processes are indirectly related to this signaling system or, even, outside this system (e.g. 5HT1A receptors, adenosine uptake, etc). This should be indicated. 5. Authors should explain the meaning of "non-medicinal cannabis products" (page 6, line 57). 6. Table 1: Children with epilepsy is poorly precise, better to mention the specific diseases: Dravet, Lennox.-Gastaut, West… 7. Some references are necessary for supporting some of the authors' comments, e.g. page 5, line 46. 8. THC should be written as Δ9-THC to differentiate from its isomer Δ8-THC.

REVIEWER	Francisco Xavier Castellanos & Paige Cervantes Department of Child and Adolescent Psychiatry, Hassenfeld Children's Hospital at NYU Langone, New York, NY F X Castellanos is on the scientific advisory board of BOL Pharma
REVIEW RETURNED	13-Nov-2019

GENERAL COMMENTS	Review questions 1. Is the research question or study objective clearly defined? Yes. However, the authors are encouraged to add an exploratory aim in the Study Objectives subsection about measuring change in severe behavior problems (SBP) pre- and post-treatment. Further, particularly if this is not an exploratory aim, the authors should consider altering the manuscript title to represent the key objectives of the study: safety and acceptability/feasibility rather than reduction in challenging behaviors. 2. Is the abstract accurate, balanced and complete? Yes. 3. Is the study design appropriate to answer the research question? The design is thorough and well thought out, especially in relation to metrics of key objectives – safety and treatment acceptability/feasibility. Several suggestions are listed below for other areas of the study design: - Phenotypic characteristics of the sample: o The intellectual disability (ID) population is a heterogeneous group, particularly when sampling across severity levels as is proposed in the current protocol. Not only will core features of ID (IQ and adaptive impairments) differ significantly between participants, but I worry that the target symptom (SBP) will also range markedly across participants. The challenging behavior profile of youth with mild ID is typically tremendously different from those with severe ID, and the features of SBP noted in the

	Introduction, such as aggression, self-injurious behavior, and banging objects are more characteristic of severe presentations of ID. A brief informant-report scale alone, like the ABC, may not pick up this distinction. o The Wechsler scales are generally thought to be suboptimal for measuring lower functioning populations, such as those with more severe forms of ID. The authors should consider using the SB-5, with also has an abbreviated IQ battery, or may want to opt for a completely nonverbal test of intelligence like the Leiter or TONI. o It is unclear why the Vineland-3 will only be administered if the participant had not had an IQ test in the previous two years. When assigning severity levels of ID, measures of adaptive function are essential. It may be preferable to bypass the Vineland-3 if both adaptive and IQ testing was completed in the previous two years but still administer the Vineland-3 if only an IQ test was administered. o The protocol reads as if autism spectrum disorder (ASD) is not exclusionary. Given the frequent overlap between ASD and ID, this makes sense clinically. However, conclusions may be difficult to draw about the ID only population if the majority of those recruited have both ASD+ID. Are there any safeguards in place to account for ratio of youth recruited with comorbid ASD? - Outcome measures: o Given the heterogeneity of challenging behavior profiles, it may be worthwhile to consider using direct observational measures. This would be relatively simple if the SPB measured is explicit (head banging) rather than representative of an internal state (irritability), and may be more clinically important and more sensitive to change than the ABC-I. Many times, school personnel and/or behavioral therapists are already recording the frequency of these behaviors and may be willing to share their data. o The authors propose a thorough evaluation of family functioning in their evaluation measures. Because of the recognized impact of SBP on parent and family wellbeing and on quality of life, this seems very appropriate. However, only one measure is included assessing participant behavior. Notably, SBP may be a functional means to getting needs met for some individuals with ID but may also represent underlying psychiatric comorbidity in others. The authors might want to consider adding a measure of psychiatric symptoms, such as the Anxiety Depression and Mood Scale (ADAMS) or perhaps a re-administration of the A-TAC. Related to potential differences in drug responsivity across behavioral functions, the authors could also add functional assessment measures. The most accurate yet resource-intensive method for determining behavioral function is the experimental functional analysis (see Danov, Tervo, Meyers, & Symons, 2012 for use of this method in an IDD drug trial [DOI: 10.1080/19315864.2011.594976]). However, the most feasible for this study may be a functional assessment rating form, such as the Question About Behavioral Function (QABF) or the Motivation Assessment Scale (MAS). o If comorbid ASD is not exclusionary, it may be worthwhile to get pre- and post-treatment estimates of ASD symptoms to evaluate change. This could be done by administering the SCQ at screening and post-treatment. If the authors believe adding an ASD measure as an exploratory outcome measure is appropriate, they may want to consider changing the SCQ to the SRS-2 to align with the tools used more frequently in ASD clinical trials (Provenzani et al., 2019; DOI: 10.1177/1362361319854641).

	o The authors should consider listing the evaluation measures on Table 2: Schedule of study visit procedures and assessments. - Drug: o The exclusion of youth currently taking antiepileptic medications may not be stringent enough given the unknowns about CBD drug-drug interactions and its significant P450 isoenzymatic profile. It may be more appropriate to exclude all youth currently taking medications metabolized primarily by CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP3A4, or CYP2D6 isoenzymes. o The dose seems high based on completed, ongoing, and proposed studies in ASD. Across the four studies examining CBD for ASD in pediatric populations listed on ClinicalTrials.Gov, the maximum dose is 10 mg/kg/day. This corresponds with the evidence suggesting lower doses of CBD may be more effective for behavioral symptoms while higher doses are better for the treatment of seizures. o The authors indicate that CBD is highly lipophilic in the Introduction. Have the authors considered how they might control for or measure the potential impact of participant diet on CBD absorption rates? 4. Are the methods described sufficiently to allow the study to be repeated? Yes. 5. Are research ethics (e.g. participant consent, ethics approval) addressed appropriately? Yes. However, determination of competency to assent is likely needed, as youth with mild and moderate ID may be able to provide assent whereas individuals with more severe presentations may not. 6. Are the outcomes clearly defined? Yes. 7. If statistics are used are they appropriate and described fully? The authors note that descriptive statistics will be used to analyze pilot study results. Because the proposed trial has a placebo group and in light of the small sample size, the authors may consider running nonparametric statistics to evaluate statistical differences in change across exploratory pre- and post-treatment measures. 8. Are the references up-to-date and appropriate? Yes. Although, in the Introduction, the authors indicate that there is “some supportive evidence emerging for its effectiveness as an adjuvant treatment to conventional AE medications for some specific epileptic syndromes.” This may be understated, as the level of evidence is substantial enough to have led to a United States FDA-approval of and DEA reclassification of CBD for Dravet and Lennox-Gestaut syndromes. 12. Are the study limitations discussed adequately? The authors could be more explicit in the Discussion about the nature of the study as a pilot to emphasize the questions left unanswered following completion of this trial and the need for further study. 13. Is the supplementary reporting complete (e.g. trial registration; funding details; CONSORT, STROBE or PRISMA checklist)? Yes.

	14. To the best of your knowledge is the paper free from concerns over publication ethics (e.g. plagiarism, redundant publication, undeclared conflicts of interest)? Yes. 15. Is the standard of written English acceptable for publication? Yes, but there are some inconsistencies in spelling and use of abbreviations (e.g., "randomized" clinical trial on page 15 and "randomised" controlled trial on page 2).

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:	
Reviewer comments	Responses
1. The major problem with the design of this clinical trial is the time for the active phase of treatment which may be too short to confirm a therapeutic effect, as it has been documented in previous trials. I would recommend a longer duration of the active phase. In addition, given the low number of patients, 5 to be treated with cannabidiol + 5 with placebo, why not using a crossed design with 50% of patients having active treatment + washout period + placebo, and the other 50% placebo + washout period + active treatment? This could be more informative with patients being the own controls during the placebo phase for the active treatment.	This is an interesting point and one that we have considered. The RCT of CBD in Dravet Syndrome reported that "the difference in favor of cannabidiol was seen in the first month of the maintenance period". This is corroborated by personal correspondence with both researchers in cannabinoids in ASD (Israel), as well as clinicians experienced in prescribing cannabinoids for this population (in Israel and Canada), where benefits were observed soon after reaching the optimal dose. Our 8 week maintenance period therefore allowed 4 weeks for treatment effects to emerge, followed by an additional 4 weeks, which corresponds with the period over which parents are required to reflect when completing the behavioural outcome questionnaire. In addition, a longer duration may negatively impact trial feasibility - in this population an unnecessarily longer treatment period may result in a larger number of treatment drop-outs for families that do not observe a positive response. Regarding the use of a crossover design, the primary goal of this study was to pilot the methodology that we intend to use in a large-scale RCT, which would be adequately powered for independent group analysis. Due to considerations such as the high lipid solubility of cannabinoids, our advice has been that a cross-over design would render interpretation of response in the second phase problematic.

2. Other question in relation with the design is to include “no previous use of cannabis” in the inclusion criteria, as it is possible that families may have treated these children with uncontrolled preparations before being recruited for the trial. This should be also confirmed with a blood analysis.	The manuscript has been updated to reflect that an exclusion criteria for this study is current medical cannabis use, or use within the 3 months prior to enrolment. (Exclusion criteria p.12). Blood cannabinoid analysis may be considered in future trials - thank you for the suggestion.
3. A comment on Epidiolex and its recent approval for infantile refractory epileptic encephalopathies is necessary, e.g. page 6, line 12.	Thank you, the manuscript has been updated. (p.6, para 1)
4. The second paragraph in page 6 would need to be rewritten. It combines the action of cannabidiol with the endocannabinoid signaling, but cannabidiol is one of the most atypical cannabinoids in relation with its activity at the classic elements of the endocannabinoid signaling. Most of its targeted proteins and processes are indirectly related to this signaling system or, even, outside this system (e.g. 5HT1A receptors, adenosine uptake, etc). This should be indicated.	Thank-you. The following sentence has been added to p.6 para 2: While THC has strong affinity for both cannabinoid receptors receptors (CB1 and CB2), CBD appears to exert its effects on the endocannabinoid system through indirect actions, and may also have activity on other neurotransmitter systems. This biochemical detail does not alter the arguments developed in the narrative – that CBD has biologically plausible potential therapeutic benefits for human behaviour.
5. Authors should explain the meaning of “non-medicinal cannabis products” (page 6, line 57).	We intended to mean cannabis products that are unregulated and not prescribed by a doctor. The manuscript has been updated to read “non-prescribed unregulated cannabis products”
6. Table 1: Children with epilepsy is poorly precise, better to mention the specific diseases: Dravet, Lennox.-Gastaut, West…	The first reference includes a heterogeneous sample of epilepsy patients, with only a small proportion diagnosed with particular epilepsy syndromes. The second reference includes Dravet syndrome (n=13), Doose syndrome (n=4), Lennox-Gastaut syndrome (n=1) and idiopathic epilepsy (n=1). The manuscript has been updated.
7. Some references are necessary for supporting some of the authors’ comments, e.g. page 5, line 46.	Thank you, the manuscript has been updated.
8. THC should be written as Δ9-THC to differentiate from its isomer Δ8-THC.	Thank you, the manuscript has been updated.

Reviewer 2:	
The authors are encouraged to add an exploratory aim in the Study Objectives subsection about measuring change in severe behavior problems (SBP) pre- and post-treatment.	Thank you, the manuscript has been updated.
The intellectual disability (ID) population is a heterogeneous group, particularly when sampling across severity levels as is proposed in the current protocol. Not only will core features of ID (IQ and adaptive impairments) differ significantly between participants, but I worry that the target symptom (SBP) will also range markedly across participants. The challenging behavior profile of youth with mild ID is typically tremendously different from those with severe ID, and the features of SBP noted in the Introduction, such as aggression, self-injurious behavior, and banging objects are more characteristic of severe presentations of ID. A brief informant-report scale alone, like the ABC, may not pick up this distinction.	The Irritability subscale of the ABC specifically assesses those target behavioural symptoms described in the Introduction (aggression, self-injurious behaviour, banging objects). It is widely used in clinical trials with this patient population. Therefore, although we agree that the behavioural profile of youth with ID can vary, this measure specifically targets the behavioural profile that we intend to treat.
The Wechsler scales are generally thought to be suboptimal for measuring lower functioning populations, such as those with more severe forms of ID. The authors should consider using the SB-5, with also has an abbreviated IQ battery, or may want to opt for a completely nonverbal test of intelligence like the Leiter or TONI.	While we acknowledge the limitations of the WASI, we are not trying to phenotype the level of Intellectual Disability or detect change in IQ over time. In this study, the WASI has been used to dichotomise the presence or absence of an ID, in order to ascertain whether inclusion criteria are met. We therefore feel that the WASI is an appropriate measure to achieve this purpose.
It is unclear why the Vineland-3 will only be administered if the participant had not had an IQ test in the previous two years. When assigning severity levels of ID, measures of adaptive function are essential. It may be preferable to bypass the Vineland-3 if both adaptive and IQ testing was completed in the previous two years but still administer the Vineland-3 if only an IQ test was administered.	This aspect of the protocol cannot be changed at this stage, however we will take this feedback under consideration when planning a fully-powered RCT.
The protocol reads as if autism spectrum disorder (ASD) is not exclusionary. Given the frequent overlap between ASD and ID, this makes sense clinically. However, conclusions may be difficult to draw about the ID only population if the majority of those recruited have both ASD+ID. Are there any	Yes we aim to recruit a pragmatic sample of youth with ID reflecting patients commonly presenting in clinical practice and in whom CBD may be considered a treatment option. ASD is difficult to diagnose in those with ID (particularly more severe ID), but as described we will attempt to identify

safeguards in place to account for ratio of youth recruited with comorbid ASD?	comorbid ASD using the A-TAC and also the SCQ, so as to describe the comorbidities present in the study sample. If an apparent signal emerges suggesting a differential response in the pilot between participants with ID only and those with comorbid ASD (or any other comorbidity) this may inform the design of the larger RCT e.g. randomisation stratification by ASD status.
Given the heterogeneity of challenging behavior profiles, it may be worthwhile to consider using direct observational measures. This would be relatively simple if the SPB measured is explicit (head banging) rather than representative of an internal state (irritability), and may be more clinically important and more sensitive to change than the ABC-I. Many times, school personnel and/or behavioral therapists are already recording the frequency of these behaviors and may be willing to share their data.	This is an interesting suggestion, however this aspect of the protocol cannot be changed at this stage.
Only one measure is included assessing participant behavior. Notably, SBP may be a functional means to getting needs met for some individuals with ID but may also represent underlying psychiatric comorbidity in others. The authors might want to consider adding a measure of psychiatric symptoms, such as the Anxiety Depression and Mood Scale (ADAMS) or perhaps a re-administration of the A-TAC. Related to potential differences in drug responsivity across behavioral functions, the authors could also add functional assessment measures. The most accurate yet resource-intensive method for determining behavioral function is the experimental functional analysis (see Danov, Tervo, Meyers, & Symons, 2012 for use of this method in an IDD drug trial [DOI: 10.1080/19315864.2011.594976]). However, the most feasible for this study may be a functional assessment rating form, such as the Question About Behavioral Function (QABF) or the Motivation Assessment Scale (MAS).	This is also an interesting suggestion, however this aspect of the protocol cannot be changed at this stage.
If comorbid ASD is not exclusionary, it may be worthwhile to get pre- and post-treatment estimates of ASD symptoms to evaluate change. This could be done by administering the SCQ at screening and post-treatment. If the authors believe adding an ASD measure as an exploratory outcome measure is appropriate, they may want to consider changing the SCQ to the SRS-2 to align	The outcome of interest in this trial is SBP, rather than ASD symptoms.

with the tools used more frequently in ASD clinical trials (Provenzani et al., 2019; DOI: 10.1177/1362361319854641).	
The authors should consider listing the evaluation measures on Table 2: Schedule of study visit procedures and assessments.	Table 2 includes 'Evaluation measures' as a group, and Table 3 expands on these with details such as item numbers, description of constructs, and rating source. We believe including all of this in Table 2 would make this table difficult to read.
The exclusion of youth currently taking antiepileptic medications may not be stringent enough given the unknowns about CBD drug-drug interactions and its significant P450 isoenzymatic profile. It may be more appropriate to exclude all youth currently taking medications metabolized primarily by CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP3A4, or CYP2D6 isoenzymes.	Again this protocol is designed for a pragmatic trial so that the findings can be meaningfully translated into clinical practice. Having said that, we have taken a somewhat more restrictive approach than the CBD in Dravet Syndrome study (Devinsky NEJM 2017), in which no patients taking concomitant medications were excluded. This is because we wanted to avoid adverse effects especially with clobazam, as our study population may be particularly sensitive to sedation which may result from impaired clobazam metabolism.
The dose seems high based on completed, ongoing, and proposed studies in ASD. Across the four studies examining CBD for ASD in pediatric populations listed on ClinicalTrials.Gov, the maximum dose is 10 mg/kg/day. This corresponds with the evidence suggesting lower doses of CBD may be more effective for behavioral symptoms while higher doses are better for the treatment of seizures.	Thank-you. In the absence of clear data to inform optimal dosing to target behaviour in this population, we opted for the higher dose demonstrated to be safe in children (with epilepsy) so that a clinical effect was not missed due to under-dosing.
The authors indicate that CBD is highly lipophilic in the Introduction. Have the authors considered how they might control for or measure the potential impact of participant diet on CBD absorption rates?	Although the intestinal absorption of some cannabis products may be increased by the ingestion of fatty foods, there is no evidence to suggest that diet is an important variable in CBD absorption when administered in an excipient oil, as is the case in this study (grapeseed oil).
Determination of competency to assent is likely needed, as youth with mild and moderate ID may be able to provide assent whereas individuals with more severe presentations may not.	The term 'assent' has no legal standing in Australia and is not recognised in the National Statement on Ethical Conduct in Human Research (2018). The Australian Paediatric Research Ethics and Governance Network (APREG) guidelines on consent reiterate that assent forms are not appropriate. Therefore, the local IRB (Royal Children's Hospital Human Research Ethics Committee) will not approve a form where the child is required to sign their assent.

	However, as per advice from our local IRB, we will involve the child/adolescent in discussions about their participation in the research and seek their agreement where possible, appropriate to their age and developmental status.
The authors note that descriptive statistics will be used to analyze pilot study results. Because the proposed trial has a placebo group and in light of the small sample size, the authors may consider running nonparametric statistics to evaluate statistical differences in change across exploratory pre- and post-treatment measures.	The study is clearly underpowered to determine efficacy and is not designed to address this question so such a test would not be appropriate.
In the Introduction, the authors indicate that there is “some supportive evidence emerging for its effectiveness as an adjuvant treatment to conventional AE medications for some specific epileptic syndromes.” This may be understated, as the level of evidence is substantial enough to have led to a United States FDA-approval of and DEA reclassification of CBD for Dravet and Lennox-Gestaut syndromes.	Thank you, the manuscript has been updated.
The authors could be more explicit in the Discussion about the nature of the study as a pilot to emphasize the questions left unanswered following completion of this trial and the need for further study.	Thank you, the manuscript has been updated.
There are some inconsistencies in spelling and use of abbreviations (e.g., “randomized” clinical trial on page 15 and “randomised” controlled trial on page 2).	Thank you, the manuscript has been updated.

VERSION 2 – REVIEW

REVIEWER	Javier Fernández-Ruiz Complutense University, Madrid, Spain
REVIEW RETURNED	04-Jan-2020

GENERAL COMMENTS	Authors have adequately addressed all comments derived from the evaluation of the first version, with the only exception of the one on the times for the active treatment. I still consider that at least a comment on this question would be of interest. This is necessary even despite authors may decide not to modify the times, but they need to explain why.

REVIEWER	Francisco Xavier Castellanos Department of Child and Adolescent Psychiatry, Hassenfeld Children's Hospital at NYU Langone, USA;

	Nathan Kline Institute for Psychiatric Research, USA Member of scientific advisory board of BOL Pharma, an Israeli company producing medicinal cannabis products; recipient of Epidiolex brand cannabidiol, provided by Greenwich Biosciences for an investigator-initiated study of children and adolescents with high-functioning autism spectrum disorder
REVIEW RETURNED	23-Dec-2019

GENERAL COMMENTS	The authors have responded constructively to reviewers' concerns and suggestions. This is a badly needed pilot effort.